The Role of A Type Lamins in Genome Organization

by
Xianrong Wong

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

December, 2015

Abstract

The nuclear lamina is composed of type V intermediate filament proteins, the lamins, that lend structural integrity to the nucleus, anchor inner nuclear membrane (INM) proteins to the nuclear envelope and have been directly implicated in maintaining genome architecture through tethering of lamina associated domains at the periphery. In mammalian cells, nuclear lamina proteins can be sub-divided into two categories, the A- (primarily consisting of the spliceoforms lamin A and lamin C from the LMNA gene) and B-type (mainly lamin B1 and lamin B2 from the LMNB1 and LMNB2 gene, respectively). While B type lamins are expressed throughout development, the expression of A-type lamins is developmentally controlled, from barely detectable expression in pluripotent stem cells and some lineage specific progenitor cells to highly abundant in many differentiated cell types. Over the years a myriad of developmental diseases, ranging from tissue anomalies, such as muscular dystrophies to the more severe and pleiotropic early aging syndromes (such as Hutchinson Gilford Progeria Syndrome), have been linked to mutations in proteins of the nuclear envelope, with many lesions occurring in the LMNA gene. While much effort in both the basic and clinical sciences continue to be invested in the investigations into the mechanisms underlying these diseases, one key question remains ignored: what are the independent roles of the A-type isoforms—are lamin A and lamin C truly functionally redundant, as most studies seem to imply?

We sought to answer this question through the lens of genome organization. Specifically, lamin A and lamin C specific shRNA mediated knockdown constructs were made to functionally test the requirement of lamin A and C in genome architecture. We show that lamin C, but not lamin A, is required both in the maintenance of lamina

associated sequences and domains at the nuclear periphery as well as the manifestation of proper functional chromosome architecture. Furthermore, we utilized the a proximity labeling technique (Biotinylation Identification or BioID) coupled with SILAC –based mass spectrometry analyses to probe *in vivo* protein interactors of lamin A and lamin C. We show that lamin C preferentially interacts with several chromatin remodelers that either have repressive roles, consistent with the interaction of this isotype with heterochromatic lamina associated domains (LADs). We also show that the localization of lamin C during mitosis differs from lamins A and B1. While lamins A and B1 localized to the newly formed nuclear periphery during telophase, lamin C exhibits a non-peripheral localization that persists through part of G1, temporally coincident with the large scale genome reconfiguration to re-establish functional architecture for homeostasis in the next cell cycle. In particular, our data support a role for lamin C in the control of higher-order chromatin structure upon mitotic exit and establishment of chromatin domains in early G1. Our results suggest that lamin C may be impacting genome architecture, in part, by directing LADs to the periphery prior to the 'locking in' of functional architecture for the next cell cycle.

Dissertation committee members:

Karen L. Reddy, PhD (Advisor/Reader)

Susan D. Michaelis (Reader)

Sean D. Taverna, PhD

Michael J. Matunis, PhD

Acknowledgments

This work would not have been possible without the generous help, support and encouragement of many people. While it is impossible to thank everyone here, I will try my best to highlight a few of them.

I would first like to thank my mentor, Karen Reddy for her support and guidance throughout my time at Johns Hopkins. My initial research focus was solely progeria centric. Due to the inaccuracies and incomplete documentation of findings in the field, I had spent the first 5 years of my graduate career trying to recapitulate the results of the field, to no avail, only to later discover that I was not alone. Karen's encouragement and support during this tough period was crucial in my decision to switch my research focus, 5 years into my PhD program.

I also want to thank Susan Michaelis for all her support, encouragement and insightful suggestions. Susan has been a close collaborator, thesis committee member and the reader of this thesis. I have really enjoyed all of our interactions and she has had a huge impact on this work.

I would also like to thank Sean Taverna for his support, encouragement and mentorship. As a member of my thesis committee and someone our lab collaborates with closely, I am fortunate for his insight and our stimulating discussions.

I would like to thank Michael Matunis, another member of my thesis committee for his input and support. Our discussions and his advice and mentorship have broadened my views on my research.

I also have to thank my lab members, Teresa Luperchio, Mohammed Heydarian, Jennifer Harr and Jevon Cutler. They have all greatly influenced my work through the

many insightful scientific discussions, physical assistance with experimentation and providing support and encouragement whenever I needed them.

To the members of the Epigenetics center, I am thankful to my colleagues in our shared lab space. Thank you all for the support and the positive influence to my personal development.

Lastly, I would like to thank my family for all their love, support and understanding. To my parents and sister, this has been a tremendously long process and I cannot thank you enough for all that you have done for me. Your encouragement and motivation served as a guiding light for me to bravely pursue what I really want in life.

Table of Contents

Abbreviations and Acronyms

2D	two-dimensional
3C	chromosome conformation capture
3D	three-dimensional
4D	four-dimensional
A	adenine
BAC	bacterial artificial chromosome
BAF	barrier to autointegration factor
bcl11a	B-cell CLL/lymphoma 11A
BioID	biotinylation identification
bp	base pair
BTB/POZ	BTB (for BR-C, ttk and bab) or POZ (for Pox virus and Zinc finger)
C	cytosine
CBS	Circular Binary Segmentation
CBX	chromobox Homolog
cDNA	complementary DNA
CFP	cyan fluorescent protein
ChIA-PET	chromatin interaction analysis by paired-end tag sequencing
ChIP	chromatin Immunoprecipitation
cLAD	common Lamina associated domain
CMV	cytomegalo virus
CO2	carbon dioxide
CTCF	CCCTC-binding factor
Dam	DNA adenine methyltransferase
DamID	DNA adenine methyltransferase Identification
DMEM	Dulbecco's Modified Eagle's medium
DNA	deoxyribonucleic acid
Dnmt3a	DNA Methyltransferase 3a
DZNep 3	deazaneplanocin
EDD	Enriched Domain Detector
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
EMD	Emerin
ESC	Embryonic Stem Cells
EZH2	enhancer of zeste 2
FB	fibroblast
FBS	fetal bovine serum
FISH	fluorescence in situ hybridization
G	guanine
G9a	histone-lysine N-methyltransferase, H3 lysine-9 specific 3

GCL	germ-cell-less
gDNA	genomic DNA
GFP	green fluorescent protein
H3	histone H3
H4	histone H4
HAT	histone acetyltransferases
HDAC	histone deacetylases
HMM	Hidden Markov Model
HMT	histone methyltransferases
HP1α	heterochromatin protein 1 alpha
Hyg	hygromycin
IF	immunofluorescence
Igh	immunoglobulin heavy chain locus
IgK	Immunoglobulin kappa locus
INM	inner nuclear membrane
IPTG	isopropyl b-D-1-thiogalactopyranoside
Ikzf1/Ik	Ikaros
K	lysine
kb	kilo base
L1/1L	LoxP site
L2β	Lap2Beta/ Lamina associated peptide-2/Beta
LacI	LacI repressor protein
lacO	lac Operator
LAD	lamina associated domain
LAP	Lamina associated peptide
LAS	lamina associated sequence
LBR	lamin B receptor
LiCl	lithium chloride
Lmn A	lamin A
LmnB1	Lamin B1
LmnC	lamin C
Mb	mega base
Me	methyl
MEF	mouse embryonic fibroblast
NaCl	sodium chloride
NPC	nuclear pore complex
NuRD	Nucleosome Remodeling Deacetylase
ONM	outer nuclear membrane
PBS	phosphate buffered saline
PcG	polycomb group
PCR	polymerase chain reaction

Pen	penicillin
PML	promyelocytic leukemia
PMT	post translational modification
PRC1	polycomb repressive complex 1
PRC2	polycomb repressive complex 2
qPCR	quantitative PCR
RAG	Recombination Activating Gene
Rb	retinoblastoma protein
RMCE	recombination mediated cassette exchange
RNA	ribonucleic acid
Seq	sequencing
SILAC	Stable isotope labeling by amino acids in cell culture
SSC	saline-sodium citrate
strep	streptomycin
T	thymine
TADs	topologically associated domains
TCIS	tagged chromosomal insertions site
TK	Thymidne Kinase
vLAD	variable lamina associated domain
YFP	yellow fluorescent protein

List of Figures

Chapter 1

Introduction

1.1 Nuclear architecture: chromatin and chromosomal domains

The dynamic organization and architecture of the nucleus contributes to regulation of gene expression during normal cellular function and throughout development, while its disorganization has been implicated in disease. In eukaryotic cells, DNA is organized into chromatin that is comprised of nucleosomes consisting of 146 bases of DNA wrapped 1.65 times around a histone protein octamer (two H2A and B heterodimers and a histone H3/H4 tetramer) (Luger et al., 1997). Chromatin (DNA plus histones with associated proteins and complexes) then serves as the substrate for virtually all nuclear events (transcription, replication, recombination, and repair) and is, therefore, highly and dynamically regulated. Nuclear organization is manifested in the arrangement of chromatin into domains of differential gene activity and compaction levels ('inactive' heterochromatin and 'active' euchromatin), individual chromosome territories, as well as subchromosomal organizational domains, such as lamina associated domains (LADs) and topologically associated domains (TADs), among others (Comings, 1980; Cremer and Cremer, 2009, 2006; Dixon et al., 2012; Guelen et al., 2008; Nora et al., 2012) . Some of these architectural domains, such as transcriptionally inactive and compact heterochromatin found in and around centromeres, near the nucleolus and at the nuclear periphery, can be observed by simply using a light micro- scope and a DNA counterstain (Boveri, 1914; Heitz, 1929)

At the molecular level, heterochromatin contains histones with specific post translational modifications (PTM) and has been shown to be (generally) more densely packed and transcriptionally 'silent' for protein coding genes [reviewed in Kouzarides, 2007]. For example, constitutive heterochromatin is enriched for histone H3 lysine 9

trimethylation (H3K9me3) and regulatory or facultative heterochromatin can be enriched in histone H3 Lysine 9 dimethylation, histone H3 lysine 27 trimethylation and/or histone H4 lysine 20 dimethylation and trimethylation (H3K9me2, H3K27me3 and H4K20me2/3, respectively). Euchromatic regions of open, transcriptionally active chromatin often carry PTMs such as acetylation on numerous lysine residues on H3 and H4 as well as H3 lysine 4 trimethylation (H3K4me3). Both H3K9me2/3 and H3K27me3 have been implicated in nuclear organization at the nuclear periphery (Guelen et al., 2008; Wen et al., 2009)

1.2 The compartment at the nuclear periphery

That chromatin is functionally organized in the nucleus is not a new idea. From Rabl to Boveri to Heitz, dynamically organized chromatin was obvious even by crude cytological measures (Boveri, 1914; Comings, 1980; Cremer and Cremer, 2009, 2006; Heitz, 1929). More recently, genome wide and cytological studies have pointed to a role for distinct nuclear compartments that are enriched in regulatory or structural proteins in both higher level genome organization and gene regulation (Mao et al., 2011). Initially, to study the three-dimensional positioning of loci relative to nuclear landmarks, such as the nuclear lamina network or peripheral zone, researchers employed fluorescence *in situ* hybridization (FISH) assays (Cremer and Cremer, 2009, 2006; Kosak et al., 2002; Kumaran and Spector, 2008; Meister et al., 2010; Reddy et al., 2008; Szczerbal et al., 2009; Williams et al., 2006). Recently, the DNA Adenine Methyltransferase Identification (DamID) technique was developed to map genome-wide chromatin contacts at the nuclear periphery, thus identifying lamina associated domains (LADs) (Vogel et al., 2007). Initial reports characterizing LADs described them as large (0.1–10

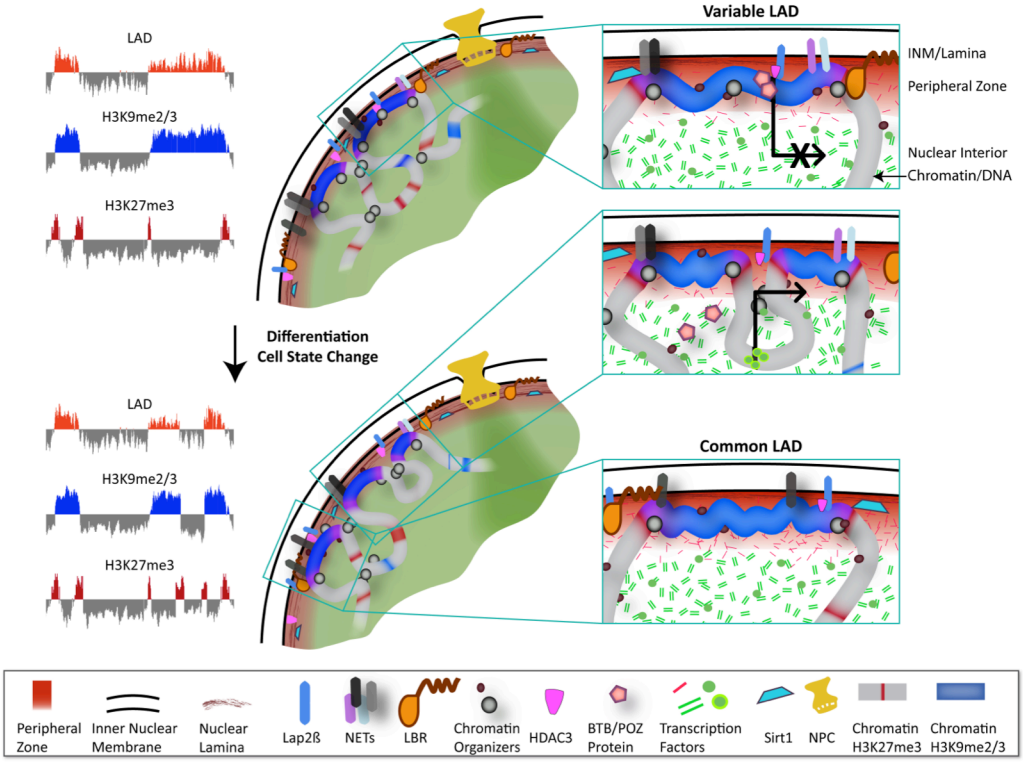
Mb) silent domains and gene 'deserts', i.e. classical constitutive heterochromatin (Guelen et al., 2008).

More recent studies have focused instead on LADs that do contain genes in order to understand the function of such organization in gene regulation (Reddy et al., 2008; Zullo et al., 2012). Interestingly, developmentally regulated genes appear to be enriched in these domains, leading to the hypothesis that LADs serve as 'facultative' heterochromatin domains during development (see below) (Meuleman et al., 2013; Peric-Hupkes et al., 2010). It is important to note that, while LADs mostly comprise inactive chromatin, there are several notable examples of lamina-proximal genes being actively transcribed (Brickner et al., 2012; Kumaran and Spector, 2008; Yao et al., 2011). These genes may interact with nuclear pore complexes, which have been shown to reside in a transcriptionally active euchromatic environment, perhaps to both facilitate transcriptional activation/re-activation and transport mRNA into the cytoplasm (Brickner et al., 2012). Nonetheless, large domains of chromatin in molecular contact with the nuclear lamina are repressed and it is these inactive, but gene-containing, LADs that comprise an active area of research to uncover the function of such domains both in chromosome architecture and gene regulation.

To this end, peripheral association of chromatin regions has been demonstrated to silence genes upon forced localization of ectopic genome sites to the inner nuclear membrane (INM) (Finlan et al., 2008; Reddy et al., 2008). These functional experiments are consistent with numerous genome-wide and cytological studies correlating cytological positioning at the lamina (Kosak et al., 2002; Meister et al., 2010; Peric-Hupkes et al., 2010; Szczerbal et al., 2009; Williams et al., 2006; Yao et al., 2011; Zullo

et.al.,2012).

Figure 1.1: Chromatin association with the nuclear periphery and scaffolding in the peripheral zone is cell type specific and dynamic through differentiation The nuclear lamina and peripheral zone facilitate gene regulation during differentiation and cell state changes through genome organization. Specifically, modifications and LAD profiles change during differentiation or cell state change (top to bottom nucleus). LAD, H3K9me2/3 (LOCK) and H3K27me3 molecular profiles for chromatin in each nucleus is depicted in panel (A). (B) cLAD regions remain at the periphery before and after cell state change (bottom inset), variable LADs move to or away from peripheral zone (red shaded region near nuclear lamina) and may be accompanied by gene activation or repression (top insets, shown to move away, with gene activation). Once at the periphery, LAD regions are retained in the peripheral zone (but are away from nuclear pore complexes – NPC in gold) and are generally overlapped by H3K9me2/3 domains (blue shading). Edges of LADs have H3K27me3 marks (red) and border regions contain both H3K27me2/3 and H3K27me3 (purple shading). H3K27me3 exists throughout the genome outside of LAD and LAD border regions (red). Proteins in the nuclear lamina and inner nuclear membrane are presumably involved in establishment of the peripheral zone (lamin A/C, lamin B1, NETs, LBR, Lap2β, Sirt1) as well as transcription factors (green and red factors throughout nuclear interior). Chromatin insulators (such as CTCF) are found both at boundaries of LADs and throughout the genome to facilitate chromosome organization. Lap2β, HDAC3 and Zbtb7b (BTB-POZ protein) have been demonstrated to be involved in scaffolding vLADs. Tissue specific factors (such as TAF3 in myoblast differentiation) are able to promote transcriptional activation of genes upon dissociation of regions from the peripheral zone that is linked with developmental progression (B, vLAD insets).



Taken together these data strongly suggest a role for the nuclear periphery in the dynamic establishment and/or maintenance of gene repression and a rapidly growing focus of research is the developmentally dynamic interaction of the genome and epigenome with the nuclear lamina and the peripheral zone (summarized in Figure 1.1). One reason for such intense focus on the nuclear periphery is that it is a unique compartment comprised of a repertoire of INM proteins and the underlying nuclear lamina. This lamina meshwork, composed of the intermediate filament coiled-coil domain proteins lamins A/C and/or B type lamins, serves as retention/docking site for the INM proteins and interfaces with the nuclear envelope (NE) and chromatin (Gruenbaum and Margalit, 2005) (Figure 1.1). Specifically, there are at least 60 unique NE proteins that contribute to the complexity of the nuclear lamina network, including INM proteins such as lamin B receptor (LBR), lamina associated peptide 2 (Lap2 also known as thymopoietin or TMPO) and emerin (EMD) (Wilson and Berk, 2010). These and other NE proteins, including proteins extending into the cytoplasm (e.g. the SUN and KASH domain proteins), interact with the nuclear lamina. In addition to other roles, many of the INM proteins, such as LBR, Emerin, and Lap2, are also able to interact with transcriptional repressors, chromatin modifiers and core histones (Guarda et al., 2009; Haraguchi et al., 2004; Holaska et al., 2003; Ye, 1997; Zullo et al., 2012). Moreover, the nuclear lamina network has been implicated directly in genome organization, epigenetic modulation and gene regulation, likely impacting both developmental and disease progression.

1.3 Proteomic diversity at the nuclear envelope

Interestingly, a number of the proteins that are resident to the peripheral zone of

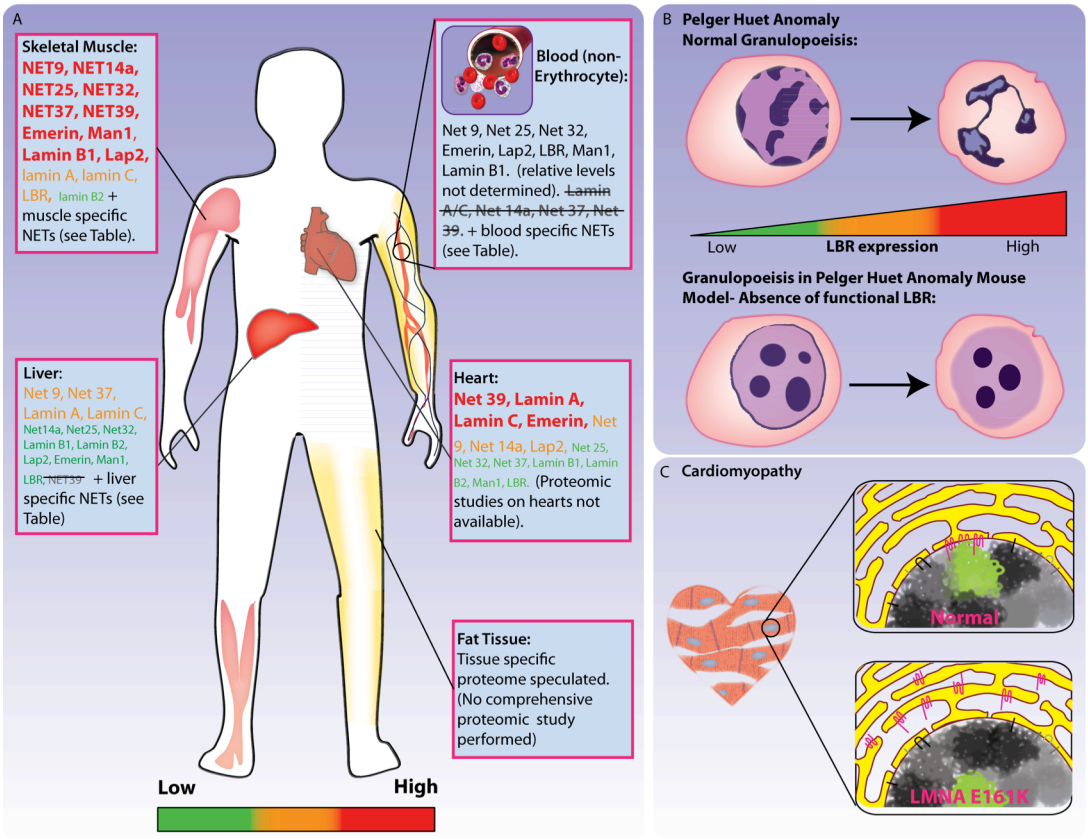


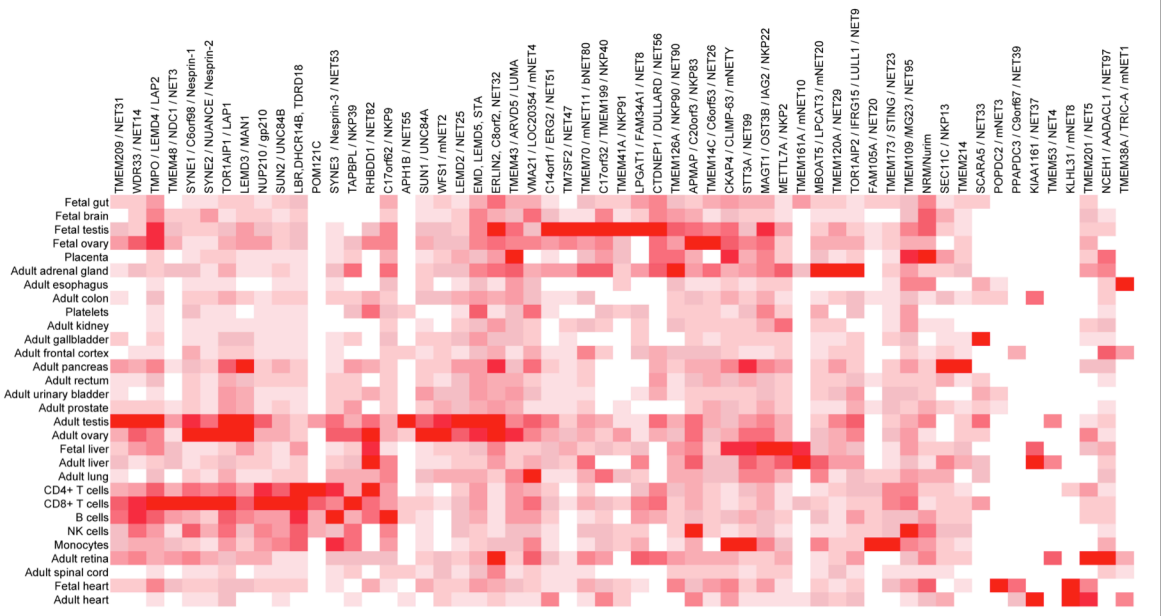
Figure 1.2. Proteomic diversity at the nuclear periphery. Shown in (A) is the inferred tissue specific expression of NETs determined by a transcriptomic analysis described in Chen et al. (Chen et al., 2006). For this representation the proteins were binned into the three expression groups for each tissue type examined; low (green), medium (orange) and high (red). It is clear, even given the small number of proteins considered here, how different the NE proteome is in divergent cell types. (B) and (C) illustrate how differential protein expression might be involved in tissue specific diseases. Depicted in (B) is the Pelger-Huet Anomaly in which the levels of LBR, a NET protein highly enriched in blood cell types and required for normal cellular physiology of granulocytes, is compromised. In normal granulopoeisis as shown in this panel, the level of LBR increases as promyelocytes differentiate into neutrophils. In neutrophils lacking functional LBR, from the *Lbr*$^{ic/ic}$ mouse model, hypo-lobulation of the nuclei and internal clumping of heterochromatin is observed. Panel (C) shows, based on speculations of existing data, how the E161K mutation in *LMNA* that should be expressed in most somatic tissues of affected individuals, displays a pathology that affects mainly the heart. In this highly speculative model, we hypothesize that the E161K mutation, which has been shown to cause disorganization of chromosome 13, leads to a disruption in normal protein-protein interactions with a cell type specific NET. One such candidate is a cardiomyocyte specific NET (BioGPS data and the human proteome map), NET39, that has been previously shown to be involved in signaling and proper chromosome positioning. Depicted in this panel, chromosome 13 (green), normally more lamina proximal in wildtype cardiomyocytes (top), is centrally positioned as a result of disrupted interactions with NET39, potentially due to mislocalization of the NET to the ER (bottom)

the nucleus — the lamins, INM proteins and newly identified Nuclear Envelope

Transmembrane proteins (NETs) — are differentially expressed (Alsheimer et al., 1999; Chen et al., 2006; Furukawa and Hotta, 1993; Furukawa et al., 1994; Jung et al., 2012; Korfali et al., 2012, 2010; Schütz et al., 2005a, 2005b; Solovei et al., 2013) (Figure 1.2). For example, the composition of the lamina meshwork has been found to be different between cell types. In brain, the lamin C isoform of the LMNA gene is widely expressed whereas the expression of the lamin A spliceoform is absent in neurons and glial cells as a consequence of selective downregulation of the pre-lamin A transcript by miR-9 (Jung et al., 2012) . Similarly, the rare lamins C2 and B3 spliceoforms, of the LMNA and LMNB2 genes respectively, have been found to be expressed specifically in mammalian spermatogenic cells, which are otherwise devoid of A-type lamins and lamin B2, both found in somatic cells (Alsheimer et al., 1999; Furukawa and Hotta, 1993; Furukawa et al., 1994; Schütz et al., 2005a, 2005b) . These spliceoforms are required for nuclear architectural changes necessary for homologous pairing in meiosis, and are therefore critical to spermatogenesis (Alsheimer et al., 1999; Furukawa and Hotta, 1993; Furukawa et al., 1994; Schütz et al., 2005a, 2005b).

More recently several proteomic studies identifying NET proteins across tissues and between cellular states have contributed to a growing list of differentially expressed proteins at the nuclear periphery (Korfali et al., 2012, 2010; Schirmer et al., 2003; Wilkie et al., 2011) (Table 1, Figure 1.2a). In these studies, a protocol to isolate and enrich for nuclear envelope proteins was developed, with a specific focus on identifying new NET proteins from specific tissues (Table 1). These experiments uncovered an impressive proteomic diversity of this cellular compartment (Figure 1.2a). In particular, a comparative study across three disparate tissues (liver, leukocytes and muscle) revealed

Figure 1.3: Tissue expression patterns of NETs. Expression patterns of NETs from Table 1 were obtained from Human Proteome Map (HPM) portal (http://humanproteomemap.org/). Using the intensities of tissue expression provided by HPM, hierarchical clustering was performed to generate a heat-map of expression, with white indicating low or no expression and red indicating high levels of expression. The tissue or cell type is indicated on the left-hand side of the diagram, while specific NETs are indicated across the top. While some NETs show limited tissue expression, such as KLHL31 and NET39, others, such as Emerin, display more pleiotropic expression.



that the majority of the 598 NETs displayed distinct expression profiles between the tissues examined, with only a modest 16% of these identified NETs being shared across all three tissue types (Korfali et al., 2012) (Figure 1.3 and Table 1 in appendix AI). Other high throughput studies have documented differential NE proteomes even in less disparate cell types, such as during T-cell activation and in addition, using mRNA expression analyses, during myogenesis (see below, NET Diversity and Nuclear Architecture). These types of studies will undoubtedly continue to identify additional novel and differential NET proteins and profiles.

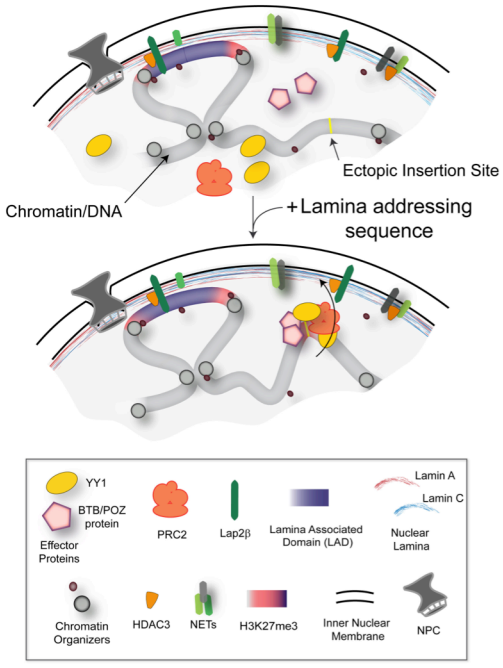
Interestingly, the draft map from the human proteome project has been made

available for browsing (http:// humanproteomemap.org/) (Kim et al., 2014). By comparing a list of confirmed NETs, it is evident that the differential NETs (Figure 1.3), previously identified in more targeted studies, also display differential expression across numerous other tissue types (Figure 1.3). It should be noted that the expression mapping in Figure 1.3 is based directly upon intensity signals from the display data from the human proteome project website (http://humanproteomemap.org/). We note that LBR seems to be enriched in hematopoietic cells, while the expression of Emerin appears to be more widespread. Additionally, Net 39 (which was previously shown to be highly expressed in skeletal muscle — a tissue type missing in the human proteome map, Figures 1.2a, 1.3) is highly expressed in both fetal and adult heart (in agreement with other expression data), but less so in most other tissues. The intersection of future NET specific data with the ongoing human proteome mapping studies will likely provide a wealth of information and much needed insights into cell type specific developmental programs and disease pathology evidenced by the Envelopathies. At the cellular level, these findings highlight the significant proteomic diversity at the nuclear periphery and the likely importance of these proteins in cell type specific functions and gene regulation/organization.

1.4 Epigenetics and topology at the peripheral zone

Early electron-microscopy data showed an obvious enrichment of heterochromatin at the peripheral zone of the nucleus. A natural hypothesis from this and other observations is that LADs will overlap canonical epigenetic characteristics of heterochromatin. Comparing genome-wide chromatin immunoprecipitation (ChIP) data

Figure 1.4: A speculative model for directed reorganization of chromatin by YY1. A cross-section of a portion of the interior of a fibroblast cell shows chromatin as a lamina associated domain (LAD, dark blue) or a non-lamina associated region (gray). Chromatin organizing factor, CTCF, and epigenetic regulators, Ying-Yang 1 (YY1) and Polycomb Repressor Complex (PRC2) are ubiquitously expressed. Cell type specific BTB/POZ type factors are present in the specified cell state (fibroblast, top). An ectopic insertion of a Lamina associated DNA fragment into an ectopic site (yellow), leads to accumulation of cell-type specific factors, such as BTB-POZ domain proteins of Zbtb7b, at the fragment containing their respective enriched binding sites. These cell type specific transcription factors may aid in YY1 recruitment or stabilization at these regions. YY1 would then recruit PRC2 enabling subsequent H3K27 tri-methylation and lamina association (bottom). There is also a requirement for H3K9me2/3 and the interaction of these two facultative heterochromatin marks (bi-valency) may be important for LAD formation and/or maintenance. Intriguingly, we also uncovered a potential role for LMNC (or LMNA/C levels) in this process.



with profiles of LADs, it has been noted that LAD borders, but not the interior of these domains, are enriched in the facultative heterochromatin histone H3 lysine 27 trimethylation (H3K27me3, Figure 1.1) (Guelen et al., 2008). Interestingly, similar ChIP experiments for the histone H3 lysine 9 di-methylation or tri-methylation (H3K9me2/3) modification shows a domain profile similar to LAD in size, distribution and position (Figure 1.1) (Wen et al., 2009). When these large organized chromatin K9 modifications (LOCKs) were compared to available LAD data, an approximately 80% overlap was noted. However, these assays were done in disparate cell types, so the extent of LAD/LOCK overlap is currently unknown. Nonetheless, it appears that these two heterochromatin marks, H3K27me3 and H3K9me2/3 LOCKs, are correlated with LADs. We note, however, that H3K27me3, unlike H3K9me2/3, is found

mostly outside of lamina proximal domains, with a higher than expected enrichment at LAD borders (Guelen et al., 2008) (Figure 1.1). Much work is focused on understanding the dynamic and causal relationship between LADs and LOCKs, that is, if LAD association is dependent upon chromatin state or vice versa (or both). Recent work from our group and others has started to bridge this gap in our understanding. In one study, a modified live-cell DamID protocol applied to a cancer cell line scheme demonstrated that knocking down G9a, the methyltransferase that deposits H3K9me3 modifications, led to a 'loosening' of LAD structure in the peripheral zone (Kind et al., 2013). A series of insightful experiments in *Caenorhabditis elegans* pointed to the role for H3K9me2 in targeting to the periphery and subsequent accumulation of H3K9me3 in these domains after becoming lamina proximal (Towbin et al., 2012). Our recent and ongoing work has demonstrated that directed association of a chromatin domain with the lamina is dependent upon chromatin state, specifically de-acetylated and enriched in H3K27me3 and H3K9me2/3 (Figure 1.4) (Harr et al., 2015; Zullo et al., 2012) . However, in embryonic stem cells (ESCs), profound differences in these domains were observed with LOCKs being largely absent and LADs displaying robust configuration (Peric-Hupkes et al., 2010; Wen et al., 2009). Interestingly, ESC genomes have also been shown to have a unique overall 3D organization (de Wit et al., 2013). Given these data, it will be important to gain a better understanding of the relationship between chromatin and LADs, especially in non-transformed mammalian cells.

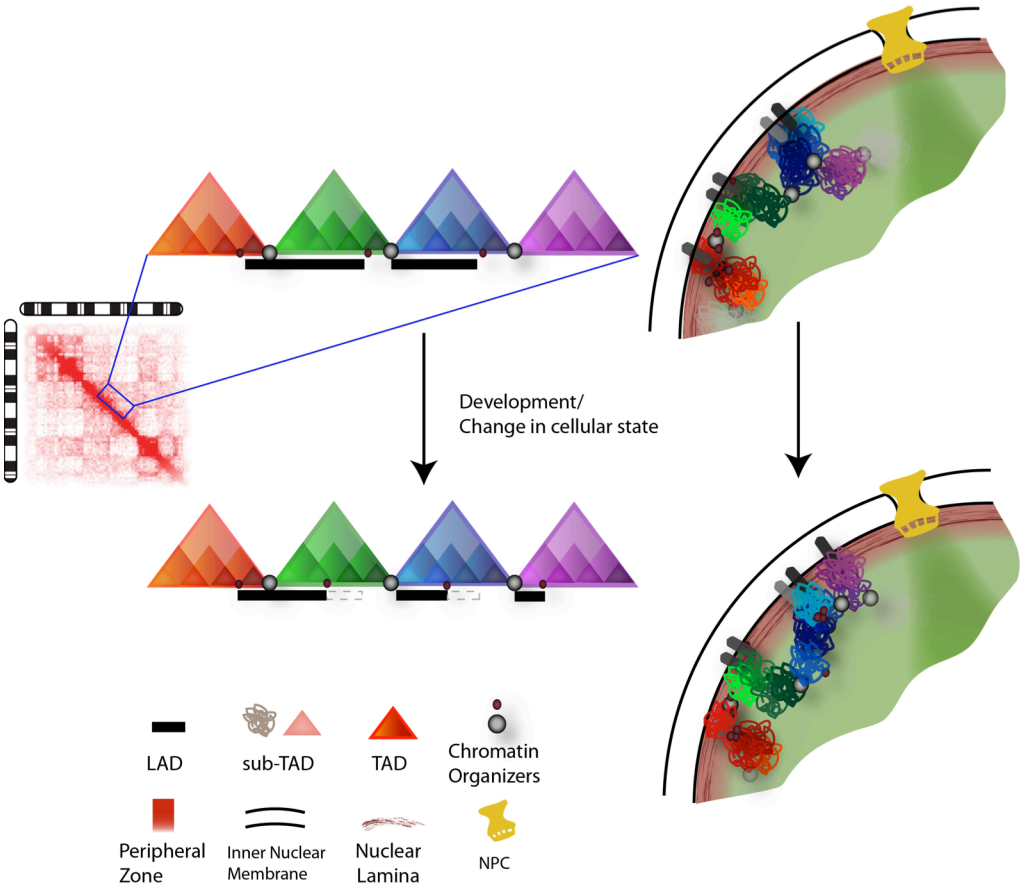
Additional analyses of genomic features in the context of LADs have revealed a few characteristics that further suggest functionality and dynamics of the peripheral compartment. Consistent with gene-poor and repressed chromatin, LINE elements and

A/T sequences are enriched in LADs, as are promoters in adjacent domains that drive transcription away from LAD border regions, perhaps contributing to the delineation of the borders (Guelen et al., 2008; Meuleman et al., 2013; Zullo et al., 2012). It is interesting to note that H3K9me2/3 domains that correlate highly with LADs have 'islands of euchromatin' within a larger LOCK domain, as proved by low H3K9me2/3 signal, and these islands overlap with transcription start sites (Wen et al., 2012). The potential functional role of these regions is unknown, but they may be indicative of regulatory regions important to LAD dynamics and/or may have a role in scaffolding.

Finally, the borders of LADs appear to be enriched in binding sites for the chromatin organizer CTCF (Figure 1.4) (Dixon et al., 2012; Guelen et al., 2008; Handoko et al., 2011; Nora et al., 2012; Zullo et al., 2012). Intriguingly, this organizer is also involved in insulator functions and higher order organization of the genome into topological domains, also known as topologically associating domains or TADs (Figure 1.2) (Dixon et al., 2012; Li et al., 2013; Nora et al., 2012; Phillips-Cremins et al., 2013; Sexton et al., 2012). TADs are large domains of the genome, detected by chromosome conformation capture derivative methods 5C or Hi-C, which preferentially interact within a single domain and exhibit very few interactions with adjacent TADs. Domainograms or association maps of interactions detected by 5C/Hi-C are aligned to the linear genome and show striking higher-level organization ('triangles' of interactions that deviate from the expected — each large triangle representing a TAD, Figure 1.5). Importantly, a study examining the correlation between genome-wide LAD borders and topological domain boundaries found significant overlap (Dixon et al., 2012). However, while the shared boundaries occurred with a much higher than expected frequency compared to chance,

many LADs crossed topological domain boundaries and vice versa (Figure 1.5). Recently, looking at maps of both TAD and LAD organization on the X chromosome, it was also noted that there were significant overlaps between these two types of domains (Nora et al., 2012) (Figure 1.5). In addition, while the authors demonstrated that overall domain organization of these topological domains are not reliant on H3K9me2 and H3K27me3 modifications (described above), there was a striking alignment with these marks. Importantly, TADs seem fairly stable between cell types and, likely, between developmental stages (Dixon et al., 2012; Nora et al., 2012; Phillips-Cremins et al., 2013). While LADs are also mostly maintained between cellular states, there appears to be a greater change in LAD organization. Importantly, a recent study highlighted the existence of a sub-TAD structure that may explain some of the higher than expected, but far from perfect correlations between LAD and TAD boundaries (Phillips-Cremins et al., 2013). Based on the existence of these sub-TAD domains (small triangles Figure 1.5), one can envision that LADs are 'built' upon this sub-TAD structure. Thus, LAD borders would fall at TAD boundaries with greater than expected frequency, but since the modular framework would really be at the sub-TAD level, the lamina-proximal domains could easily fall across two TADs and even contain borders within the TAD (Figure 1.2). Obviously, how LADs and TADs (or substructures within) are related is still an area of intense investigation.

Figure 1.5: Topological domain structure is reconfigured at the sub-TAD level during differentiation or changes in cell state (A) Model depicting TAD structure and relationship to LAD profiles during differentiation/cell state change. LAD domains are modularly reorganized based on sub-TAD organization, and the two domains share boundaries. Chromatin insulators are integral for TAD domain boundaries, as well as LAD domain boundaries. (B) Representation of nuclear architecture based on TAD and LAD structural maps from panel (A). As LADs are lost or gained, specific sub-TADs move away or toward the lamina/peripheral zone (red shading, green shading indicates nuclear interior).



1.5 Functional and dynamic organization of LADs in development and cellular function

To be clear, it is unknown whether LADs identified by DamID or other means are actually scaffolded at the INM/lamina or whether they represent a population sequestered to a more peripheral zone (Figure 1.1). Recent live cell studies in transformed cells demonstrated that LADs are dramatically and dynamically repositioned after transiting the cell cycle, even well away from lamina- proximal positions (Kind et al., 2013). It is unclear, given experimental limitations (overexpression of the Dam-LMNB1 fusion protein and the use of a highly transformed cell line, for example), whether these long range and apparently stochastic movements are indeed a facet of normal LAD organization and dynamics. However, these studies uncovered an interesting phenomenon occurring at the nuclear periphery — an apparent stochasticism and dynamics within what we are calling the 'peripheral zone' (Figure 1.1). The idea for such a zone is supported by studies in developing myoblasts which identified the endogenous Myod1 gene as localized to the periphery and lowly expressed (Yao et al., 2011). Moreover, the locus was retained in the peripheral region until activation/differentiation. Specific exclusion of the core transcription factor TAF3 away from the zone at the periphery was demonstrated to be important for the relative inactivity of the MyoD1 locus in early myogenesis and H3K9me3 was demonstrated to be enriched in this same zone. In addition, TAF3 was demonstrated to require interactions with H3K4me3 through its PHD domain to sequester it away from the lamina proximal zone, indicating that this active chromatin modification is excluded as well. Interestingly, artificially tethering a Myod1 promoter transgene to the nuclear periphery results in total transcriptional repression of the transgene, raising the

provocative possibility that protein-lamina tethering and localization to the peripheral zone both serve to regulate gene activity, but to differing degrees. Numerous other studies have shown interactions of chromatin modifiers and scaffolders at the periphery, for example HDAC3 interactions with Lap2b, HP1 with LBR and BAF (Barrier-to-integration factor) with numerous INM proteins and chromatin(Berk et al., 2013; Olins et al., 2010; Zullo et al., 2012) . As demonstrated by the regulation of the MyoD1 gene and as will be discussed (in the next section) for the differential roles of LBR, compositional changes in the peripheral zone or nuclear lamina network have effects on gene regulation and chromatin scaffolding.

This raises the question of how a genic region might be recruited or directed to, or away from, the peripheral zone. While evidence supports that gene activation and movement away from the peripheral zone are coupled, it is currently unclear whether activation precedes movement away from the repressive domain, and is a target of study (Almouzni et al., 2013; Kumaran and Spector, 2008; Peric-Hupkes et al., 2010). No definitive 'LAD' zipcode has yet been identified across all LADs and it is important to note that, since cLADs and vLADs exhibit different responses to developmental cues, not all LADs are equal. It will therefore be important to analyze vLADs specifically to uncover sequences and mechanisms that allow for reorganization of genic regions in a cell type specific manner. As an example, a recent study focused on the Immuno-globulin Heavy Chain and Cyp3a loci (which are both lamina-proximal in murine fibroblasts but nucleoplasmic in pro-B cells and hepatocytes, respectively) uncovered dispersed Lamina Associated Sequences (LASs) that can direct repositioning of ectopic regions to the nuclear periphery when randomly integrated into the genome (Zullo et al.,

2012). The LASs identified were enriched in GA di- nucleotides, which is striking given that LADs are relatively A/T rich. It is tempting to speculate that such LASs may be important for establishing peripheral compartmentalization of genomic regions during cellular differentiation. It is, however, still unclear how decisions pertaining to cell type specific LADs are being made given an (almost) identical genome across cell types of an organism. It will be necessary to uncover more such LAS elements from vLADs and to determine whether such sequences are truly sufficient and what proteins mediate reorganization into a de novo LAD.

The above mentioned study also uncovered roles for a cell type specific transcription factor (Zbtb7b), a histone deacetylase (HDAC3) and an INM protein (LAP2b) in mediating de novo interactions with the nuclear lamina (Figure 1.1). Zbtb7b (also known as ThPok) encodes a BTB-POZ domain protein that can act as a transcriptional repressor or activator and recognizes $GA_{(n)}$ sequences (Matharu et al., 2010; Melnick et al., 2002). The transcriptional repressor mGCL (murine Germ-Cell-Less), another BTB-POZ domain protein, also interacts with Lap2b and Emerin in the INM, prompting the intriguing possibility that this large family of differentially and developmentally expressed proteins may play a role in LAD architecture and function (Kimura et al., 2003). In addition, the BTB-POZ domain has been shown to recruit PRC2 (Polycomb Repressive Complex — responsible for H3K27me3) to chromatin, leading to the compelling possibility that targeting of LASs by these proteins leads to H3K27me3as well as association with the peripheral zone (Boulay et al., 2012). Interestingly, in Drosophila neurons, the neural competence factor hunchback (Hb) becomes repressed and lamina-proximal concordant with down-regulation of distal antenna (Dan) (Kohwi et

al., 2011). Dan contains a 'pipsqueak' GAGA DNA binding motif that competes with BTB-POZ domain proteins for its target sites, and enforced expression of Dan prevents the lamina- proximal positioning and repression of Hb (Kohwi et al., 2013; Lehmann et al., 1998). LAS containing regions, therefore, likely recruit specific factors (e.g. Zbtb7b or Dan) that in turn either promote (Zbtb7b) or inhibit (Dan) association with the peripheral zone (Figure 1.1). In addition, these cell type specific transcription factors could recruit chromatin remodelers for the induction of large scale chromatin changes concurrent with either localization to or away from the peripheral zone. Moreover, epigenetic state may be the overriding factor in the 'decision' to be relegated to the peripheral zone and LASs with specific protein partners may merely act as a molecular switch for inducing chromatin signatures that drive to the peripheral zone. In either case, it appears that some sort of combinatorial information (protein binding or epigenetic state, for instance) on top of the LASs is needed to discriminate between domains that remain nucleoplasmic and those that are destined to become LADs in a cell type specific manner. It is likely that more cell type specific factors (both the previously mentioned NETs and transcriptional regulators) will be identified in regulating the dynamic LAD landscape, both in development and disease.

1.6 NET diversity and nuclear architecture

How can one invoke a relatively large domain 'scaffolded' at the nuclear lamina that is also dynamic? Clearly, both genome organization and protein composition at the nuclear periphery change with differentiation and cellular state. As discussed above, the mechanism of positioning and maintenance of large chromosomal regions at the nuclear periphery is a subject of intense investigation. Such subchromosomal domain positioning

is likely to rely on a combination of chromatin modifications, DNA sequence, DNA or chromatin binding proteins and proteins of the INM/lamina. To this end, a handful of studies have looked into the contribution of the changing proteome at the nuclear periphery to elucidate the impact of differential genome organization in disparate cell types.

In attempts to test the ability of several NETs in influencing genome organization, two separate studies screened a fraction of NETs, identified in the above-mentioned proteomic studies, for their ability to affect chromosome localization and chromatin compaction (leukocyte activation study) using both overexpression and knockdown strategies (Korfali et al., 2010; Zuleger et al., 2013). As discussed previously, in leukocytes, the organization of the genome and the proteome at the nuclear periphery changes dramatically upon phytohemagglutinin mediated T-cell activation (Figure 1.1). These differences were suggested to be responsible for the dissipation and reorganization of the massive amounts of dense heterochromatin normally found at the periphery in unstimulated T-cells (Figure 1.1b) (Korfali et al., 2010). To test this hypothesis, several NETs identified to be differentially expressed in unstimulated as well as activated T-cells were overexpressed in HT1080 fibrosarcoma cells. Altered locus localization was measured using a LacI-GFP fusion protein binding to lacO repeats inserted into chromosome five. In this experiment, STT3A and TAPBPL overexpression favored the localization of chromosome 5 to the nuclear periphery, which is consistent with their expression status and genome organizational changes between the two cell states. Additionally, the NET IAG2 was determined to have strong effects on chromatin compaction in an overexpression assay. In a second study by the same laboratory, NETs

were screened for their ability to influence genome architecture in overexpression studies in HT1080 fibrosarcoma cell lines as well as with shRNA mediated knockdowns in relevant cell types that express the NET under investigation. Several NETs, including NET39, were identified as important for chromosome positioning. Interestingly, NET39 was one of six NETs identified to be upregulated, by mRNA expression arrays, during myoblast differentiation and was subsequently shown to be important in the negative feedback inhibition of IGF-II via its interactions with mTOR (Chen et al., 2006; Liu et al., 2009). Intriguingly, the NETs identified that were capable of regulating chromosome and locus localization were found to be expressed in only a restricted subset of tissues by western blot, suggesting that cell type specific and differential genome organization might be, at some levels, regulated by differential NET expression (Table 1, Supplemental Table 1 for identification of NETs that affect chromo- some positioning or organization).

Additionally, lamin B receptor (LBR), a well-known differentially expressed NET previously implicated in neutrophil development and nuclear organization(Gaines et al., 2008; Hirano et al., 2012; Shultz et al., 2003; Subramanian et al., 2012), has been implicated along with A type lamins in heterochromatin recruitment or scaffolding at the nuclear periphery (Hirano et al., 2012; Solovei et al., 2013). Specifically, the inter-relationship between LBR and A type lamins was examined during development and in terminally differentiated cells (Solovei et al., 2013). Normally, heterochromatin is seen to be scaffolded at the NE in most mammalian cell types. However, post-mitotic cells depleted for LBR and A-type lamins display an inverted chromatin configuration with heterochromatin residing in clumps in the nucleoplasm. The authors note that this is

reminiscent of rod cells in nocturnal animals that are naturally devoid of LBR and A type lamins (Figure 1.1f). Centrally localized heterochromatin was suggested to act as a 'lens' that focuses light, facilitating vision under low light conditions (Solovei et al., 2009). We also note that this is highly reminiscent of the architectural changes observed in granulocytes in Pelger–Huet anomaly, which lack functional LBR (see below and Figure 1.2). Interestingly, while LBR is sufficient for repositioning heterochromatin to the nuclear periphery in rod cells, ectopic lamin C expression did not prevent the inversion of heterochromatin location (Solovei et al., 2013). This suggests that A-type lamins scaffold heterochromatin at the nuclear periphery by means of mediators, such as NETs, that are anchored onto the lamina meshwork. Consistent with the above findings, LBR has been shown separately to be able to bind specific chromatin modifications (H4K20me2) and to promote chromatin condensation (Hirano et al., 2012). In this study, a stepwise model for heterochromatin formation by LBR was proposed: LBR first binds to chromatin via the chromatin modification H4K20me2, then clustering of LBR allows high affinity tethering of chromatin at the nuclear periphery. Thereafter, epigenetic modifiers such as MeCP2 and HP1 mediate heterochromatin formation and repression of the tethered domain. Intriguingly, the importance of LBR levels and function has been implicated in other developmental systems, most notably, granulopoiesis (Figure 1.1c). During the developmental progression of a myeloid progenitor to a neutrophil, LBR is dramatically upregulated and is required for the nuclear lobulation characteristic of this cell type, as well as the localization of condensed chromatin to the nuclear periphery (Figure 1.1c). It is tempting to speculate that other NET-lamina interactions also lead to differential organization and regulation of the genome.

Obviously, many more studies are needed to identify other NETs that play a role in chromosome organization, especially in cell types where there are possible links between genome organization, development and disease (Figure 1.6). For example, during adipogenesis, the nucleus has been documented to undergo dramatic architectural genome reorganization, including repositioning of entire chromosome territories along with adipogenesis specific genes (Figure 1.1, 1.6) (Kuroda et al., 2004; Szczerbal et al., 2009)). However, it is unknown at this time how different the nuclear envelope proteome is between pre-adipocytes and adipocytes. Thus, we can only speculate that this observed genome organization during development may be linked, in part, to a change in the nuclear envelope proteome (Figure 1.1d). Investigations into adipocyte specific NETs would also lend insight into specific laminopathies in which the fat tissue is dysregulated, for instance familial partial lipodystrophy (FPLD) and Hutchinson–Gilford Progeria Syndrome (HGPS). Such speculations could be extended to numerous cell types where similar genome regulation and architectural changes have been observed in the nucleus as part of the normal developmental program or disease pathology.

1.7 Diseases of the nuclear periphery

Given the involvement of the nuclear periphery in a multitude of cellular processes, it is perhaps not surprising that proteins at the nuclear periphery have been implicated in a range of developmental diseases. It is thus not far-fetched to speculate that differential NET composition will impact the severity, penetrance and tissue specificity of these diseases. These Nuclear Envelopathies are sometimes referred to as laminopathies, although the latter usually refers specifically to mutations in lamin genes (Nagano and Arahata, 2000; Worman, 2012) . Prior to high throughput explorations of

the chromatin and proteomic environment of the nuclear envelope and INM (through proteomic, transcriptomic, epigenomic and DamID approaches), it was confusing how seemingly ubiquitous and important regulatory proteins resulted in such highly tissue specific disorders. Naturally, there has been the supposition that differential protein–protein interactions likely played a role. High throughput transcriptomic and proteomic approaches, characterizing differential expression of NE specific proteins in divergent cell types, have however provided an opportunity to uncover new protein partners and mechanisms behind tissue specific manifestations of the Envelopathies. Importantly, these studies have dramatically expanded the list of proteins that reside at the nuclear periphery and some of these have previously been demonstrated to have other cellular functions. It is not difficult to speculate that tissue specific proteomic signatures at the nuclear periphery may contribute to the cell and tissue specificity of envelopathic diseases.

Envelopathies can arise from mutations in tissue restricted NETs or from mutations in more pleiotropically expressed NETs or lamina proteins. In the case of the former, disease manifestation in a specific tissue type makes sense, since the protein functions are normally restricted to a specific subset of cells. An example of this would be Pelger–Huet anomaly (discussed in greater depth below), an autosomal dominant disease, caused by mutations in LBR that generally lead to loss of sterol reductase function, although it is currently unknown if these diseases are driven solely by loss of function of sterol-reductase activity (Subramanian et al., 2012). In the case of the latter, however, disease manifestation may be tissue restricted due to a number of reasons, including differential mechanical stress, extracellular environment and, we emphasize,

differential NE proteome composition of the affected cell types. Examples of Envelopathies in this class include HGPS, FPLD, and autosomal dominant EDMD (Emery Dreifuss Muscular Dystrophy), which are all caused by mutations in the widely expressed LMNA gene, and X-linked EDMD, which is caused by mutations in the widely expressed ED gene (Figure 1.3).

As mentioned above, mutations in LBR give rise to PHA and related diseases. Typically, the disease causing mutations lead to a loss or perturbation of the sterol reductase domain in LBR, although it is still unclear if loss of sterol reductase activity is the sole cause of disease phenotype. During granulopoeisis, as promyelocytes (or earlier precursors) differentiate into granulocytes, the expression of LBR increases and is correlated with the degree of segmentation or hyper-lobulation of the granulocyte nuclei, chromatin condensation at the nuclear lamina, and neutrophil development (Figure 1.1c, 2b). In PHA, neutrophils from patients fail to become hyper-lobulated but instead display more hypo-lobulated nuclei (Figure 1.2b). Using mice that lack a functional LBR ($Lbr^{ic/ic}$), it was demonstrated that neutrophils derived from $Lbr^{ic/ic}$ promeylocytes also failed to undergo hyper-lobulation (Zwerger et al., 2008) (Figure 1.2b). The inability of PHA neutrophil nuclei to gain hyper-lobulation has been attributed to the compromised sterol reductase activity of LBR (Subramanian et al., 2012). Interestingly, in addition the nuclear shape dysregulation (hypo-lobulation), improper chromatin organization is evident in neutrophils from $Lbr^{ic/ic}$ mice. Specifically, this anomaly in chromatin organization manifests in a failure to scaffold heterochromatin at the nuclear periphery and a notable aggregation of heterochromatin in the nucleoplasm (Shultz et al., 2003). This is reminiscent of some of the more recent studies in rod cells, as discussed above,

and is consistent with the supposed role LBR plays in scaffolding heterochromatin to the nuclear periphery (Hirano et al., 2012; Solovei et al., 2013)(Figure 1.1c, 2b). Given the high expression levels of LBR in neutrophils, and other blood cell types, and the dependence of LBR on proper functionality of this cell type, it is understandable that the neutrophils are the first affected cells upon slight perturbations to the levels of functional LBR, as seen in PHA heterozygosity. It is of special importance to note that LBR is expressed in other tissues, albeit at lower levels (Figure 1.3). Total abrogation of LBR expression, on the other hand, has been documented to affect several other tissues, including skeletal anomalies (Greeenberg Dysplasia) in addition to abnormalities to neutrophils [77,78]

Contrary to the more straightforward explanation of PHA and associated diseases where tissue specificity can be explained by the relative abundance of the affected protein in a given tissue, other tissue specific diseases of the nuclear periphery are less straightforward to elucidate since they arise from mutations in more pleiotropically expressed proteins. It is in these cases where a better understanding of the differential NET proteome is important for understanding disease pathology. Specifically, the very widely expressed LMNA gene has been implicated in both systemic as well as more tissue specific diseases (Bonne et al., 1999; De Sandre-Giovannoli et al., 2002; Eriksson et al., 2003; Fatkin et al., 1999; Holt et al., 2001; Lin and Worman, 1993; Magracheva et al., 2009; Muchir et al., 2000; Novelli et al., 2002; Raffaele Di Barletta et al., 2000). A few of the diseases caused by mutations in LMNA are discussed below.

Perhaps the most widely known laminopathy is Hutchinson–Gilford Progeria Syndrome (HGPS) which is a severe premature aging defect that affects multiple tissues

and most commonly results from the silent mutation 1824 C>T. This mutation activates a cryptic splice site in the LMNA gene, leading to an in-frame 50 amino acid deletion in the mutant protein, commonly referred to as progerin (Eriksson et al., 2003). HGPS is an autosomal dominant disease and ectopic expression of progerin, in some cells, results in cellular aberrations that include the loss of peripheral heterochromatin, loss of H3K27me3 and H3K9me3 PTMs (globally), 'blebby' looking nuclei, down-regulation of Lap2b and lamin B1 and, at late stages, inversion of the RAN gradient necessary for nuclear import and loss of general genome organization, as measured by Hi-C (Goldman et al., 2004; Kelley et al., 2011; Liu et al., 2011; McCord et al., 2013; Scaffidi and Misteli, 2006, 2005; Shumaker et al., 2006; Zhang et al., 2011). Interestingly, as previously mentioned, in neural and glial cells only the lamin C isoform of the LMNA gene is expressed due to suppression by a miRNA. Consequently, progerin is not produced in these cells, sparing HGPS patients neural or cognitive defects (Jung et al., 2012). Importantly, one of the most affected tissues in HGPS is the vascular endothelium and as a consequence patients most often die from strokes and myocardial infarctions (Eriksson et al., 2003). We should note, however, that other tissue types, including adipose tissue and skeletal muscle, are also affected by progerin expression. While it is clear that progerin expression leads to the disease phenotypes, it is unclear why specific tissues, such as the vasculature, are so strikingly affected.

Contrary to the pleiotropic disease HGPS, other mutations in LMNA lead to much more obvious tissue/ cell type restricted diseases, such as lipodystrophies (FPLD), muscular dystrophies (EDMD) and cardiomyopathies (Bonne et al., 1999; De Sandre-Giovannoli et al., 2002; Fatkin et al., 1999; Holt et al., 2001; Lin and Worman, 1993;

Magracheva et al., 2009; Muchir et al., 2000; Raffaele Di Barletta et al., 2000). How different mutations in a widely expressed gene result in such varying clinical phenotypes is only just beginning to be understood. It is likely, again, that the cell type specific NET proteome is at least in part responsible for the tissue specificity of these diseases. As an example to illustrate this hypothesis, we highlight the E161K mutation in LMNA that causes dilated cardiomyopathy with early atrial fibrillation (Mewborn et al., 2010). Gene expression analysis has been carried out using explanted hearts from patients with dilated cardiomyopathy resulting from the E161K mutation or from a non-LMNA linked cause (control) (Mewborn et al., 2010). Results from this analysis showed that chromosome 13 had a higher than expected fraction of misexpressed genes in the heart and fibroblasts of the E161K patient. Moreover, E161K patient fibroblasts consistently had a more centrally disposed chromosome 13 as compared to fibroblasts from patients with other LMNA mutations that affects striated muscles and fibroblasts from a normal individual. Interestingly, NET39 (PPAPDC3) is 16 times as highly expressed in human heart compared to all other tissues (BioGPS data and human proteome map, Figure 1.3) and as mentioned previously, has been shown to influence the radial position of chromosome 13, favoring a more peripheral disposition (Zuleger et al., 2013). These correlative data prompt us to speculate that the E161K mutation affects the proper localization or function of NET39 and that the resulting perturbed NET39 functions lead to both dysregulation of the genome and signaling (both functions attributed to NET39 in other cell types) (Figure 1.2c). This example underscores the need for the characterization of many more tissues in order to understand cell/tissue specific developmental processes and pathology.

Unfortunately, the proteome of the nuclear periphery of many other tissues and cell

types remain unexplored. For instance, in FPLD, a disease resulting from specific mutations in LMNA that affects fat tissues and causes metabolic disruptions, understanding the changing/ differential proteome during adipogenesis (discussed above) or even simply characterizing the protein composition at the NE of adipocytes has the potential to help identify accessory causes to the manifestation of this disease. Similarly, there may be unsuspected accessory proteins (in addition to LUMA and proteins of the LINC complex) involved in the observed comorbidity of muscular dystrophy and cardiomyopathy in patients with LMNA and EMD mutations (Liang et al., 2011; Meinke et al., 2011). Profiles of skeletal muscle and heart nuclear envelope proteins will likely provide candidates (such as NET39/PPAPDC3, discovered in just such a study, Figure 1.3) for future studies on proteins complicit in tissue restricted laminopathic disease phenotypes.

Similar to the paucity of cell or tissue types profiled for the NE proteome, an in depth knowledge of the nuclear functions of many of these identified NET proteins, which without doubt will aid mechanistic studies of disease causation, is still lacking. Moreover, in spite of the innumerable investigations of functions of even the widely known NE proteins, eg LBR and lamin A/C among others, many unanswered questions remain. This deficit in knowledge is best illustrated using the A type lamins, one of the more widely studied lamina proteins, as an example. In mammalian cells, the major A type lamin isoforms are LmnA and LmnC. Apart from the unique C terminal ends (6 unique residues on the C terminal end of LmnC and 77 residues on the C terminal end of mature LmnA), these two proteins are otherwise, identical. Henceforth, many studies investigating the functions of A type lamins implicitly assume that lamins A and C are

functionally redundant and are studied collectively. However, the question of whether lamins A and C are in fact redundant, has not been queried. This, unfortunately, is an important question because, it will allow the discrimination of different pathways or network of proteins that become affected in the different laminopathies having mutations common to both lamins A and C. In addition, given the importance of the nuclear periphery, especially lamins A/C (and LBR), in regulating genome function, the interface between the lamina and chromatin is just beginning to be studied.

Henceforth, it appears that we still have a long way to go in our understanding of how differential proteomes at the nuclear periphery affect the dynamics of LADs and influence genome architecture and how the dysregulation of these interdependent modules causes diseases. Complementing the use of 'omics' tools for the profiling of the epigenetic landscape, LADs, the transcriptome and the proteome of the nuclear envelop with in depth functional characterization of proteins at the nuclear envelope is thus needed to advance our understanding of the ongoing dynamics at the nuclear periphery and provide insights to treating the laminopathies.

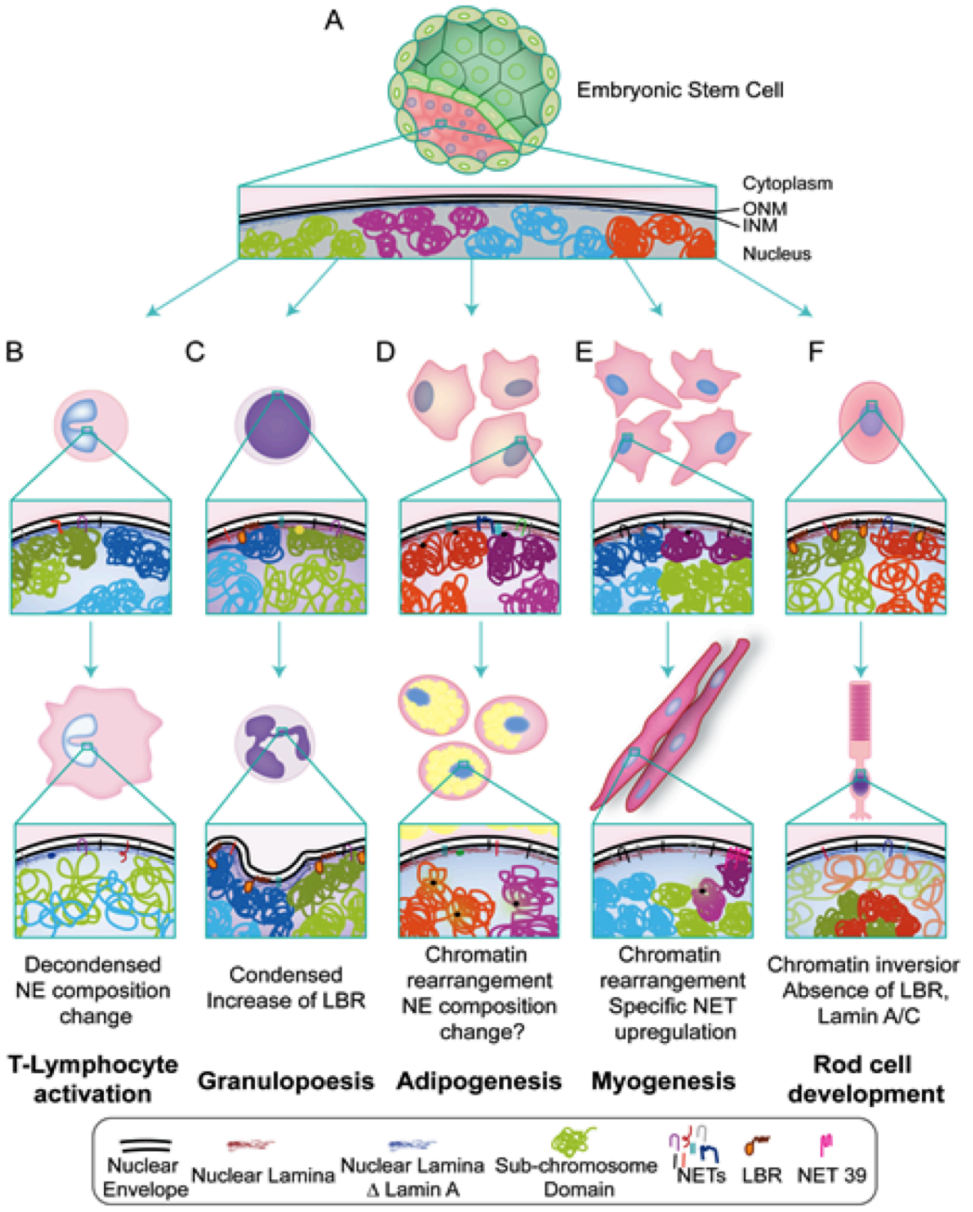


Figure 1.6: Tissue specific nuclear architecture and NETs observed during development/ differentiation.

Embryonic stem cells (ES cells, A), which differentiate into all cell types, have a unique nuclear architecture as determined by the chromosome conformation capture method, Hi-C. NET composition of the nuclear envelope of embryonic stem cells is largely unknown, but they are devoid of (or only lowly-express) lamin A (blue lamina indicates a lamina meshwork devoid of Lamin A) which may contribute to their unique organization.

Figure 1.6 (continued): Tissue specific nuclear architecture and NETs observed during development/ differentiation.
We also highlight several different cell types and subsequent nuclear architecture changes that occur upon stimulation or developmental progression. We highlight changes that occur during T cell activation (B), granulopoiesis (C), adipogenesis (D), myogenesis (E), and rod cell development (F). During T-cell activation (B), the composition of the nuclear envelope changes and the activated T-cell (B, bottom) displays generally decondensed chromatin and a notable loss of heterochromatin at the nuclear periphery. During granulopoesis (C) as promyelocytes (C, top) differentiate into neutrophils (C, bottom) there is a characteristic increase in Lamin B Receptor(LBR) and a concordant decrease in Lamin A, which results in increased heterochromatin at the nuclear periphery and the characteristic hyper-lobulated nuclear morphology of the neutrophil (C, bottom). Adipogenesis (fat cell differentiation, D) also demonstrates nuclear architecture changes, and a number of differentiation specific genes have been shown to move away from the nuclear periphery and into the nuclear interior upon activation (from a LAD to a non-LAD, D, bottom, genes depicted as black dots on chromatin domains). We speculate that this process is accompanied by a nuclear envelope proteome change. Myogenesis, or differentiation of myoblasts into myotubes and ultimately muscle tissue (E), also exhibits a change in the NET proteome, specifically the expression of NET39, which is important for proper positioning of chromosome 5 and 12. *MyoD*, a gene encoding a key regulator of myogenesis, is also observed to assume a more central location within the myotube nucleus (black region of chromatin, E) and this movement corresponds with upregulation of *MyoD*. During rod cell differentiation and maturation (F) nuclear architecture dramatically changes with the displacement of heterochromatin to the center of the nucleus in an inverted conformation. This reorganization is concurrent with the loss of LBR at the nuclear periphery, in a process that appears to be the opposite of what occurs in granulopoesis (C). Each distinct cell state has a unique nuclear architecture and nuclear envelope proteome that are both important for proper development and differentiation.

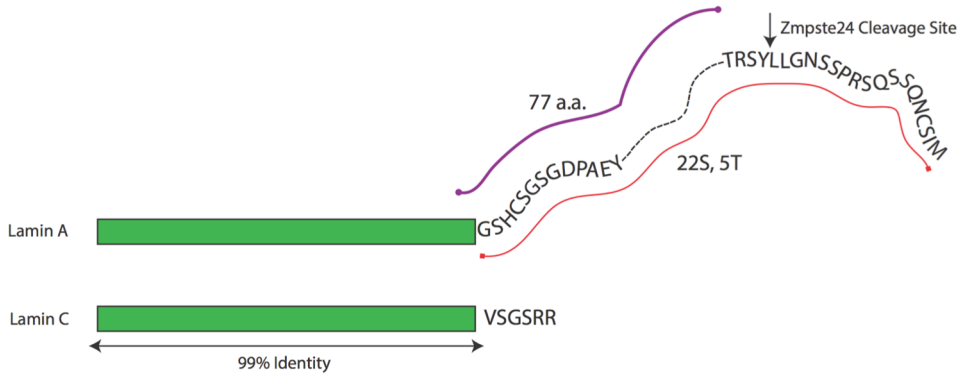


Figure 1.7: Commonalities and differences between the LmnA and LmnC protein. The identical portion of both proteins is shown by the green bar (99% identity). Lamin A has an approximately 97 residue tail which is then cleaved by Zmpste24 to yield a 77 residue tail. In this unique C terminal tail of lamin A, there are 27 potential sites for regulation by phosphorylation and O-GlcNac modification. Lamin C, in contrast has a 6 unique amino acid tail at the C terminus.

Chapter 2

An adaptation of the DamID Protocol to Deep Sequencing

2.1 Introduction

The dynamic organization and architecture of the nucleus contributes to regulation of gene expression during normal cellular function, in development, and its disorganization has been implicated in disease. That chromatin is organized in the nucleus is not a new idea. From Rabl to Boveri to Heitz, dynamically organized chromatin was obvious even by crude cytological measures (Boveri, 1914; Comings, 1980; Heitz, 1929). More recently, genome wide and advanced cytological studies have pointed to a role for distinct nuclear compartments, such as the nuclear lamina, that are enriched in regulatory or structural proteins in both higher level genome organization and gene regulation (Mao et al., 2011).

Nuclear organization is manifested in the arrangement of chromatin into domains of relatively open and transcriptionally active domains (euchromatin) and more compacted regions that are more inactive and/or contain structural elements (centromeres and telomeres) of the genome.. In addition, chromosomes, which contain both heterochromatin and euchromatin sub-domains, are themselves constrained into chromosome territories and subchromosomal organizational domains, such as lamina associated domains (LADs) and topologically associated domains (TADs), among others (Comings, 1980; Cremer and Cremer, 2009, 2006; Dixon et al., 2012; Guelen et al., 2008a; Nora et al., 2012). A particular focus in recent years has been the study of LADs, which are very apparent heterochromatic domains juxtaposed to the nuclear envelope by cytology, which have been more recently identified by a genomics technology called DNA adenosine methylation identification (DamID) (Guelen et al., 2008a; Peric-Hupkes et al., 2010a; Zullo et al., 2012a). Using DamID, several groups have identified that 30-

40% of the genome is organized, to some level, into these regions. Intriguingly, there are apparently different types of LADs--those that harbor developmentally regulated genes, and change their LAD status upon differentiation, and those that appear more constitutive (Luperchio et al., 2014; Meuleman et al., 2013; Peric-Hupkes et al., 2010b). How the domains at the nuclear lamina are regulated and the inter-relationship between regulation at the lamina, the changing nuclear envelope proteome, the epigenome and dysregulation of this region by mutations in nuclear envelope proteins is an active area of research. The nuclear periphery is a unique compartment comprised of a repertoire of INM proteins and the underlying nuclear lamina. This lamina meshwork, composed of the nuclear coiled-coil domain proteins lamin A/C and B type lamins, serves as retention/docking sites for the inner nuclear membrane (INM) proteins and bridges thee nuclear envelope (NE), chromatin and the cytoplasm (Gruenbaum and Margalit, 2005) (Figure 2.1). Specifically, there are at least 60 unique NE proteins that contribute to the complexity of the nuclear lamina network, including INM proteins such as lamin B receptor (LBR), lamina associated peptide 2 (Lap2 also known as thymopoietin or TMPO) and emerin (EMD) (Wilson and Berk, 2010). These and other NE proteins, including proteins extending into the cytoplasm (e.g. the SUN and KASH domain proteins), interact with the nuclear lamina and connect the nuclear envelope to both to the underlying chromatin/nucleoplasm and the cytoplasm. In addition to other roles, many of the INM proteins, such as LBR, Emerin, and Lap2, have been shown to interact with transcriptional repressors, chromatin modifiers and even modified core histones (Guarda et al., 2009; Haraguchi et al., 2004; Holaska et al., 2003; Ye, 1997; Zullo et al., 2012a).

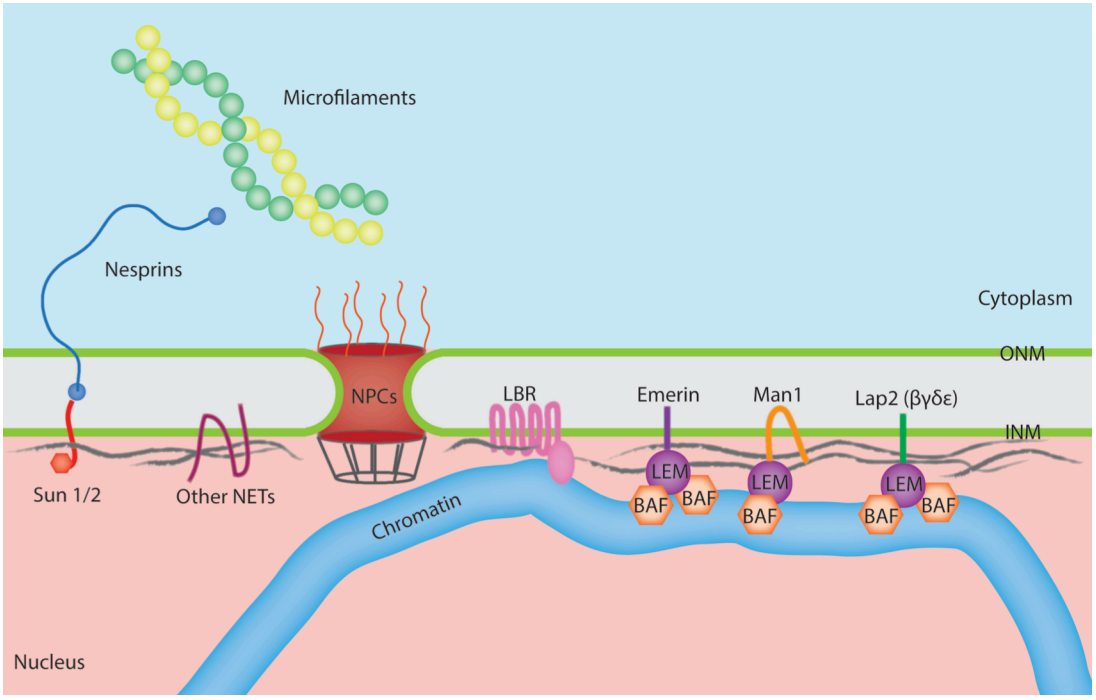


Figure 2.1. Schematic illustration of the nuclear periphery. The nuclear periphery consist of a dual membrane nuclear envelope, studded with nuclear pore complexes (NPCs). The inner nuclear membrane (INM) comprise of a plethora of proteins, held in place by the underlying lamina, that have diverse cellular functions, including chromatin tethering, epigenetic modulation and cytoplasmic-nuclear communication – among others.

There are numerous mutations in both lamin proteins and INM proteins that lead to disease and, therefore, understanding the relationship between this unique compartment and chromatin regulation, in particular, is of great import.

Initially, to study the three-dimensional positioning of loci relative to nuclear landmarks, such as the nuclear lamina network or peripheral zone, researchers employed fluorescence in situ hybridization (FISH) assay(Cremer and Cremer, 2009, 2006; Kosak et al., 2002a; Kumaran and Spector, 2008; Meister et al., 2010a; Reddy et al., 2008a; Szczerbal et al., 2009; Williams et al., 2006)s. In these studies, several developmentally regulated loci were shown, by 3D-immunoFISH (FISH with immunofluorescence, mostly

to the lamina) to be associated with the lamina in developmental stages prior to their activation and to move away from the lamina upon (or shortly prior) to activation. On the heels of these observations, several groups undertook functional studies using forced recruitment of ectopic loci to the nuclear periphery to determine the functional consequences of association with the INM/lamina (Finlan et al., 2008; Kumaran and Spector, 2008; Reddy et al., 2008a). These studies demonstrated, to varying degrees, that genes were silenced upon association with the INM/lamina. In addition, similar experiments demonstrated that genomic regions at the lamina can be activated with strong transcriptional activators (Chuang et al., 2006). These functional experiments are consistent with numerous genome-wide and cytological studies correlating cytological positioning at the lamina with low transcriptional output, but also with the ability to become activated when strongly induced (Chuang et al., 2006; Kosak et al., 2002; Meister et al., 2010; Peric-Hupkes et al., 2010; Szczerbal et al., 2009; Williams et al., 2006; Yao et al., 2011; Zullo et al., 2012). Taken together these data strongly suggest a role for the nuclear periphery in the dynamic establishment and/or maintenance of gene repression. Moreover, the nuclear lamina network has been implicated directly in genome organization, epigenetic modulation and gene regulation, likely impacting both developmental and disease progression.

More recently, the DNA Adenine Methyltransferase Identification (DamID) technique was developed to map genome-wide chromatin contacts at the nuclear periphery thus identifying lamina associated domains (LADs) (Guelen et al., 2008).

This technique has been used to query LAD signatures across multiple cell types, namely embryonic stem cells (ES cells), neural progenitor cells (NPCs), astrocytes (ASCs) and

embryonic fibroblasts (MEFs) and pro-B cells (Harr et al., 2015; Peric-Hupkes et al., 2010). Comparing LAD maps throughout the neural development scheme, it was noticed that genes within domains that gain association with the nuclear lamina mostly become repressed and genes within domains that lost peripheral contact become activated or poised for expression, in agreement with the earlier cytological studies (Peric-Hupkes et al., 2010; Kosak et al., 2002; Meister et al., 2010; Peric-Hupkes et al., 2010; Szczerbal et al., 2009; Williams et al., 2006; Yao et al., 2011).. This type of dynamic subnuclear localization of key developmental genes coupled with variable LAD maps throughout a developmental lineage provides evidence for a functional type of LAD, as opposed to the original generally 'silent' definition of LADs (Guelen et al., 2008). LADs have since been described to comprise of at least two different subclasses: 'constitutive LADs' and 'facultative LADs' (Meuleman et al., 2013). While much evidence exists for subclassing of LADs, we feel these terms are a bit misleading given the paucity of cell types studied and the still ambiguous role of individual LADs and changing LAD landscapes. Between cell types, we find evidence of LAD regions that are common (cLADs) and those that are variable (vLADs), which may play either structural and/or regulatory roles (Peric-Hupkes et al., 2010; Reddy et al., 2008; Zullo et al., 2012 unpublished data) (Figure 1.1, introduction). Nonetheless, an interesting analysis of these LAD subtypes have been documented. While developmentally important genes are enriched in variable LADs, cLADs have been reported to have high A/T content and to be highly enriched in specific classes of repetitive elements such as LINEs – among others, giving rise to speculations of potential functional attributes these repeat elements might present, in LAD formation. This is especially intriguing given the described regulatory nature of several repeat

elements in instructing genome function both in prokaryotes and eukaryotes. These include, but are not limited to, the origin of replication (OriC) in Escherichia Coli. that instructs DNA replication, telomeres' involvement in the formation of the synaptonemal complex in meiosis, kinetochore formation on centromeres during mitosis and the gypsy transposable elements as insulators to enhancer functions in Drosophila Melanogaster (Bass, 2003; Chen and Corces, 2001; Fuller et al., 1984; Westhorpe and Straight, 2013). However, these analyses were performed on data obtained from tiled arrays which are mostly repeat-masked. As a consequence, the enrichment of repeat elements in cLADs is, at best, inferred from the larger surrounding domains identified bioinformatically. A deep sequencing approach in which repeat elements are retained in the dataset will be needed to definitively ascribe prospective functional roles or domain enrichments to the different classes of repeat elements.

Intriguingly, a hallmark of cancer is the uncontrolled expression of repeat elements such as satellite sequences and transposable elements. Furthermore, the aberrant expression of transposons can alter the structure and regulation of genes when excised and reinserted into protein coding sequence or regulatory elements governing gene expression (Chénais, 2013). Moreover, such transposition events have been shown to introduce not just interrupting transposon sequences, but also duplication of genomic DNA. this occurs through a process called 3' transduction in which transcription of the L1 element extends beyond the L1 repeat resulting in the introduction of genetic content (in addition to the transposon itself) from the original transposon location to the site of insertion (Goodier, 2000). While it remains unclear whether the uncontrolled expression of repeat elements are drivers of oncogenesis or merely a consequence of transformation,

the deleterious effects posed by their expression certainly lead to a downward spiral in terms of genome stability and hence cellular malfunction. Henceforth, the ability to map the enrichment of these repeat elements in specific nuclear compartments and their mislocalization in a diseased state provides for a rigorous diagnostic tool for cancer prognosis.

In this chapter, we describe an adaptation of the DamID protocol to deep sequencing using the Illumina Hiseq2000 platform. We emphasize that the downstream bioinformatics analysis was performed to yield the maximal resolution of the technique and the utility of sequencing technologies to potentially yield information of nuclear localization of repeat elements. In addition, we provide strong evidence that the bioinformatics tools used to determine LAD organization are vital to understanding these domains and we therefore, directly compare some published algorithms.

2.2 Determining lamina associated domains using a high throughput DNA sequencing adaptation of DamID

In this study, we sought to identify all genomic sequences associated the nuclear lamina by adapting the DNA adenine methyltransferase identification (DamID) protocol for deep sequencing. This technique relies on expressing a fusion of a protein of interest, in this case the nuclear lamina protein LaminB1, with Dam methyltransferase from E.*coli*. in cultures cells. The Dam portion of the fusion is able to modify adenine residues with a methyl group at the N6 position of the purine, resulting in $G^{m6}ATC$ tracts in regions of the genome where the fusion protein interacts. Because the fusion is restricted to the nuclear lamina, the GATC modifications preferentially occur in these regions. A

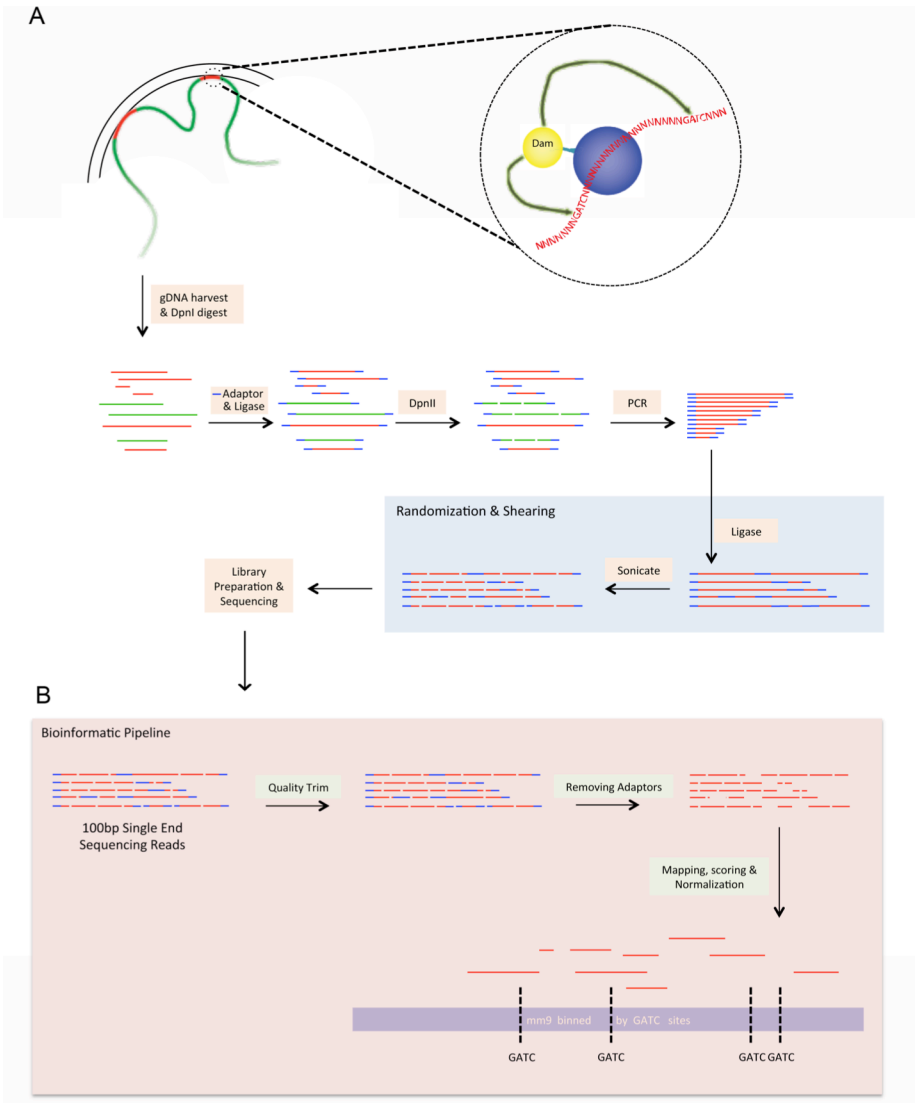


Figure 2.2. Dam-ID library processing and analysis are outlined in (A) and (B) respectively. (A) Cells expressing Dam or Dam-LmnB1 are lysed and harvested for genomic DNA. Genomic DNA was then digested with the *Dpn*I restriction enzyme that recognizes G^{m}ATC. Double stranded adaptors, in blue, with overhangs are then ligated onto the *Dpn*I digested DNA. Post ligation, *Dpn*II was used to remove fragments that contain internal unmethylated GATCs to remove genomic fragments that are not in molecular proximity to the nuclear periphery, in green. This is followed by a ligation mediated PCR that specifically amplifies genomic fragments having peripheral association, in red. Current deep sequencing technologies require a library size of between 100 to 500 bp for appropriate clustering on the flow cell. Since Dam-ID relies on restriction enzymes, each genomic fragment that gets amplified is from a specific genomic locus and the thus DamID amplified material is non-randomized. Direct sonication of the DamID material may, as a result, cause the loss of smaller fragments. The DamID material was thus randomized by ligating the fragments together, followed by sonication to the required size for clustering. These resulting fragments were then subjected to sequencing library preparation. (B) 100bp single end sequencing reads are quality controlled by trimming the ends of reads that fail certain criteria based on quality scores. The reads are then removed of DamID adaptor sequences resulting from the randomization (ligation) procedure. Each resulting (red) fragment greater than 25nucleotides in length is then mapped back to the genome build by bowtie.

control Dam only construct is introduced in a separate cell culture dish to normalize for random Dam/DNA interactions. For this study, lenitviral particles harboring genes for Dam-LmnB1 or Dam were made and used to transduce primary mouse embryonic fibroblasts and immortalized fibroblasts (Guelen et al., 2008). The Dam containing genes were allowed to be minimally expressed for at least 48 hours before genomic DNA extraction and $G^{m6}ATC$ specific amplification, using the $G^{m6}ATC$ specific enzyme DpnI followed by ligation mediated PCR. (Fig 2).

In order to subject the amplified material to deep sequencing technologies on the Illumina platforms (in this case the Illumina HiSeq 2000), DNA library sizes typically have to be in the range of 100 to 500bp for efficient clustering on the flow cell and maximal sequencing read yield. Importantly, unlike Chromatin Immunoprecipitation, the reliance of DamID on the restriction enzyme DpnI dictates that each amplified size fragment derives from specific genomic loci. Consequently, direct shearing of the DamID amplified material would lead to bias in terms of genomic representation of fragments as a result of sonication. We therefore first randomized our amplified material by ligation (Figure 2.2). This randomization step was followed by a sonication step to yield DNA fragments ranging between 100-300 bp which were then subjected to standard DNA sequencing library preparation for 100 b.p. single end reads (NEB; catalog no. E7370S). For all DamID deep sequencing experiments, material from cells expressing the Dam gene was used as a parallel control, to normalize for local DNA accessibility and sequencing bias.

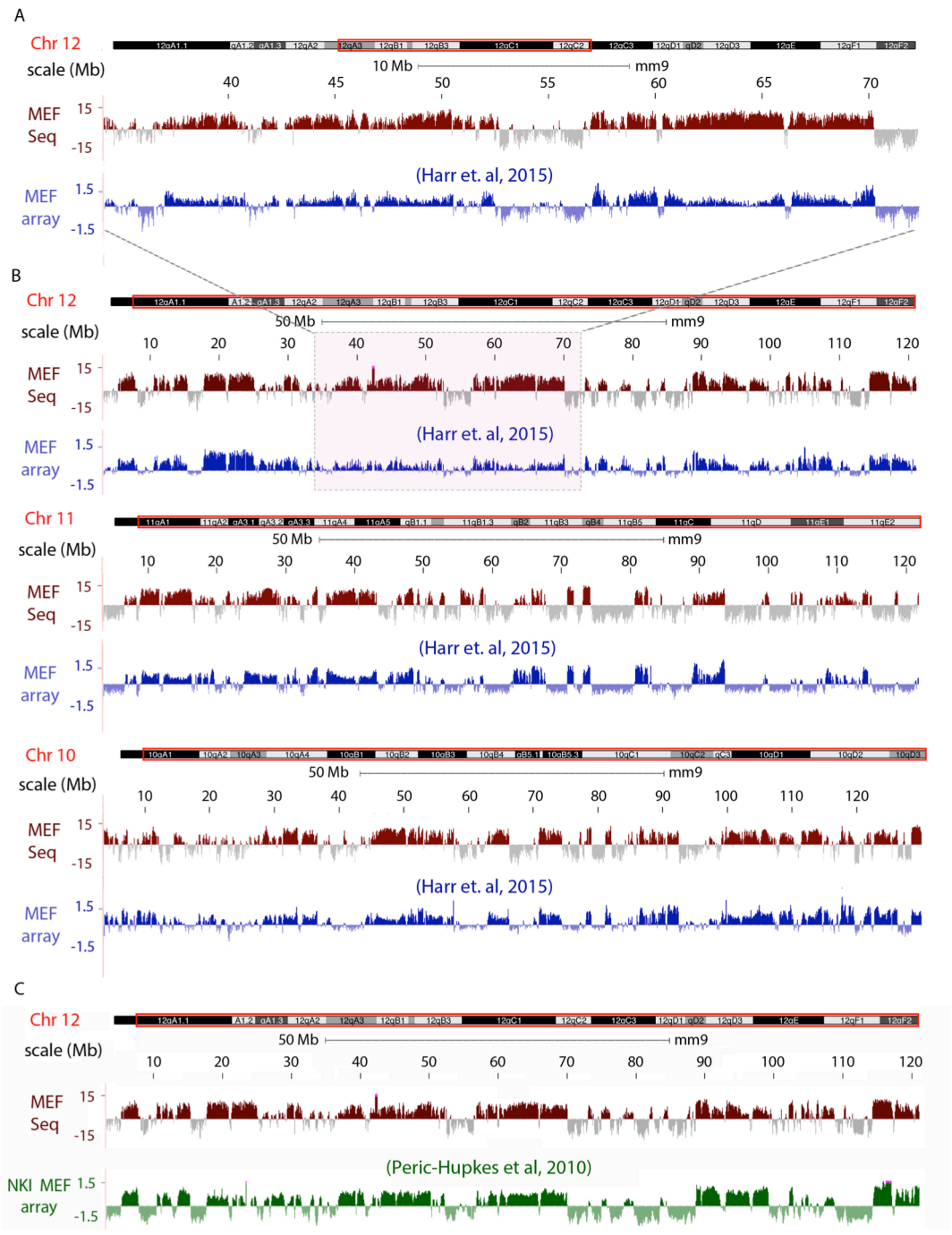


Figure 2.3. Validation of DamID-Seq pipeline. Comparison between data from Dam-ID array and Dam-ID sequencing was visualized in UCSC genome browser. The data is displayed in the form of a histogram on a $\log_2$ scale in which a darker shade represents a positive log2 score for LmnB1/Dam. (A) shows the trace of a region in chr12 comparing MEF-array (blue) with MEF-sequencing (maroon) data. (B) shows the comparison of MEF-array (blue) with MEF-sequencing (maroon) data for chromosomes 10-12. (C) shows the comparison of our DamID-seq data (maroon) with the NKI MEF array data (Green).

2.3 Reads Mapping and Lamina Associated Domain Segmentation

As a consequence of our randomization procedure as described in the previous section, some of our reads will have internal DamID primer sequences which are then subjected to removal via an in-house written script (Fig 2, Appendix AII)). In this process, genomic sequences that flank the DamID primer sequence were treated as individual reads which are then subjected to mapping using bowtie. For all mapping purposes, a filter of sequencing read length greater than 25 nucleotides were set.

Although DamID amplified material has mostly been applied to tiled arrays there are a few studies that have implemented massively parallel deep sequencing approaches (Carl and Russell, 2015; Guelen et al., 2008b; Harr et al., 2015; Peric-Hupkes et al., 2010b; Wu and Yao, 2013; Zullo et al., 2012b). In a recent report outlining DamID sequencing, the authors binned the mouse genome into contiguous 2kb windows for scoring (Wu and Yao, 2013). In another recent approach, using haploid cells and single - cell DamID-seq to identify allele specific and cell-to-cell variation in LAD organization, the bins were even larger, at 100Kb resolution (Kind et al., 2015). In our approach, we reasoned that since the DamID technique relies on the methylation of adenines in GATCs in the genome, followed by identification of interacting domains with *Dpn*I, which cuts these sites, binning the genome using the GATC sites as delimiters of bin size (hereafter called *DpnI* bins) would yield the maximum resolution allowed by the technique. For each *Dpn*I bin, a score is assigned based on the number of reads that overlap the bin, normalized by the bin size (for larger bins we expect a larger number of reads) and the total number of mapped reads. This normalization was done for both Dam-Lamin B1 and Dam only samples. These scores were then used to obtain a $\log_2$ ratio ($\log_2$ [Dam-

LaminB1 score/Dam only score]) by normalizing the scores from the Dam-LmnB1 experiment to the Dam only control, yielding genome wide *Dpn*I scores for each DamID sequencing experiment, in which a positive number indicates positive association with the nuclear lamina. Displaying these *Dpn*I score data on the UCSC genome browser we show that our genome-wide profiles are remarkably similar to the original DamID chip experiments and more recent DamID – sequencing approaches. Upon comparing these DamID-seq data with previously published array data and our own unpublished array data, the profiles are remarkably similar ((Harr et al., 2015; Peric-Hupkes et al., 2010b), Figure 2.3).

As previously documented, lamina associated domains (LADs), are large domains of DNA in close proximity to the nuclear periphery. These domains range in size from 0.1 to 10Mb. It is often of interest to segment or identify intervals of LADs so that other genome wide datasets, such as transcription factor binding sites regions of particular histone modifications, DNA methylation, etc., –, can be bioinformatically intersected with the identified LADs to investigate how other genomic features relate to these domains. To identify LAD regions as intervals of genomic coordinates (i.e. segmentation), we used an implementation of a circular binary segmentation (CBS) algorithm as described previously (Harr et al., 2015; Zullo et al., 2012b). CBS is an algorithm that was originally used to identify genome copy number variation (CNV) by recursively segmenting the genome until those segments that have probe/bin distributions different from the neighboring segments have been identified. The original use of this algorithm was to identify duplications of genomic DNA and, hence, is ideal at detecting large domains in which the signal changes from neighboring regions (see methods).

Using this approach, we identified 5064 LADs in our primary MEF dataset. We first confirmed that the CBS algorithm, which we originally used on array based data (Harr et al.), was compatible with sequencing datasets using *Dpn*I bins (Figure 2.4A). We next compared the LADs we found with previously documented MEF LADs from array based DamID followed by HMM-mediated LAD identification (Peric-Hupkes et al., 2010). LAD data from the NKI MEFs using Hidden Markov Model based algorithm was taken directly from the published results and database (Peric-Hupkes et al., 2010). Remarkably, more than 90% of NKI MEF LADs coincide with our MEF DamID-seq LADs with a coverage of 90-100% (Figure 2.4B, left panel). However, , only 1763 out of 5064 of our DamID-seq LADs are covered by the NKI MEF data(4B, right panel). To determine where this discrepancy came from, we next used the LADs identified from either our data (DamID-seq or DamID-array followed by CBS segmentation) or NKI MEFs (DamID-array followed by HMM-mediated LAD identification) and plotted these against the raw data from each experiment. Specifically, we plotted genome wide averages of lamin B1 interaction intensity (raw data) extending from -2.5kb to 2.5kb of LAD borders (as determined by the LAD segmentation algorithms) with the aid of the R based Genomation package. In all cases, the identified borders clearly separated non-LAD from LAD domains in a sharp, clearly delimited transition, suggesting that both methods and algorithms are accurately identifying LAD domains (4C). However, the HMM algorithm model appeared to miss many LADs from array based data, so we next asked whether this was because of the different array design, as the arrays used in the

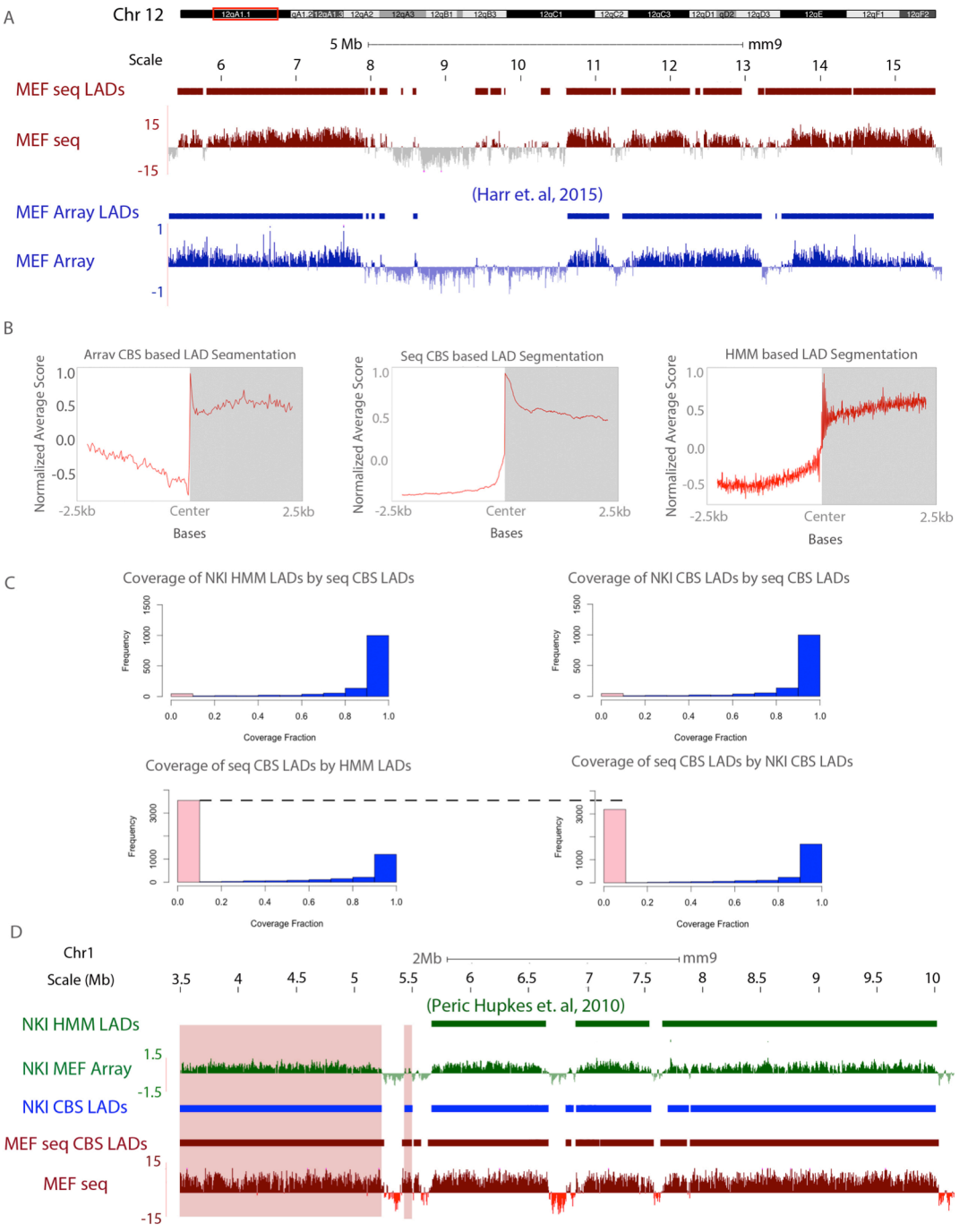


Figure 2.4. Comparison between CBS and HMM based segmentation. (A) Visual traces showing $\log_2$ ratios of Dam-Lamin B1/Dam and identified LADs for a region in chr12 for MEF-array (blue) and MEF-sequencing (maroon) data. (B) Plots of the transition of average lamin B1 occupancy across all LAD borders (white: outside LADs, gray: inside LADs). (C) Histograms of LAD coverage (top: coverage of NKI MEF array – HMM based lads by Sequencing – CBS LADs (left) and vice versa (right), bottom: coverage of NKI MEF array-CBS lads by Sequencing-CBS LADs (left) and vice versa (right). Pink bars highlight LADs with zero coverage. (D) Traces of log2 score for LmnB1/Dam from NKI MEF array (green) and MEF seq (maroon) with their respective LADs in the same colors above the traces of log2 score. Blue trace shows the LADs obtained in this region when the NKI MEF array data was input into the CBS algorithm.

NKI MEF study were custom designed and not as tightly tiled as our arrays, or the setup/design of the algorithm itself. Our DamID-array and DamID-seq data show robust correlation in coverage both ways (data not shown). Upon visual inspection of the LAD calls for the NKI MEFs, it was clear that the algorithm missed regions that were clearly enriched for LaminB1 interaction (Figure 2.4D, green bars versus green histogram). Therefore we took the original NKI MEF array data $\log_2$ ratios (Figure 2.4D, green histograms) and ran this data through our CBS pipeline. The resulting CBS identified NKI MEF LADs (blue bar below green histograms) more closely match our MEF sequencing data (red bar, red histograms), indicating that the CBS algorithm is able to rescue many of these 'lost' LADs. (Figure 2.4C)

We next compared the LADs we found with previously documented MEF LADs from DamID-seq followed by SICER-mediated LAD identification (sLADs) (Wu and Yao, 2013). Again, while over 90% of sLADs identified overlapped with our MEF LADs, only 2112 out of 5064 of our DamID-seq MEF LADs are covered by LADs identified by SICER in a MEF DamID-seq approach (Figure 2.5A).. While the SICER identified sLADs were indeed over large domains of positive association with the periphery, these bioinformatically identified LADs show false internal truncations within what are clearly regions of high LAD signal, resulting in small peaks being identified instead of larger domains (Figure 2.5B). Such failure to identify the actual LAD breadth and more contiguous interactions would be very detrimental to downstream approaches intersecting with other large genome wide datasets, such as transcription data (RNA-seq)

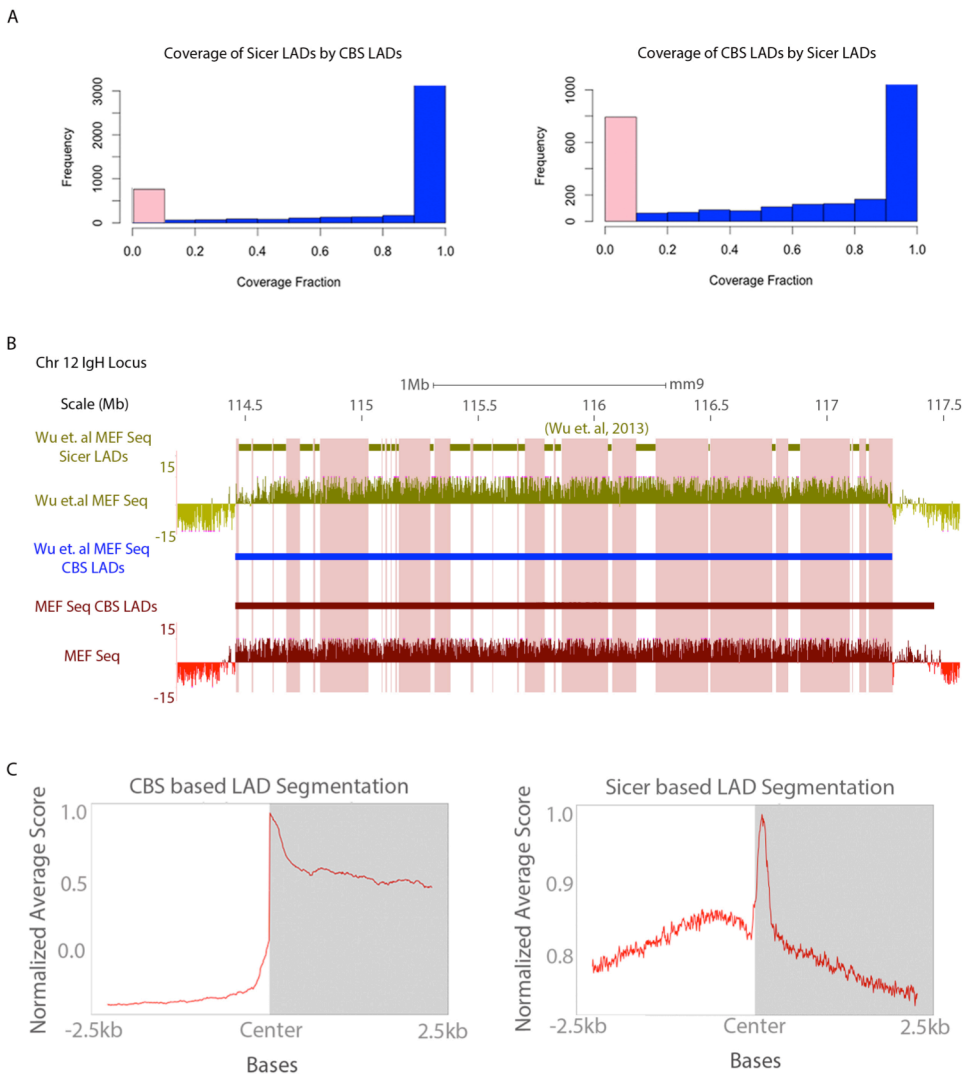


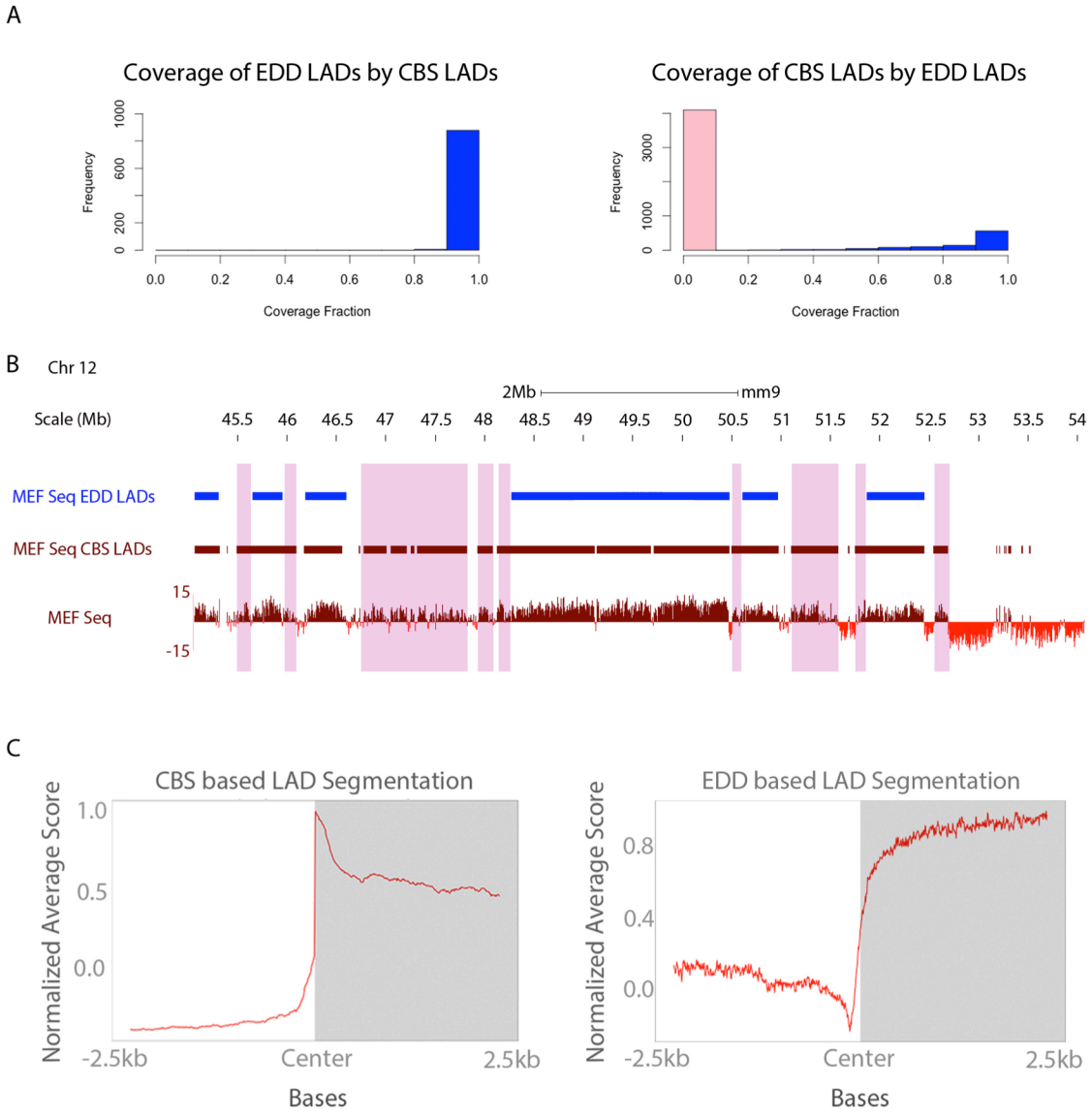
Figure 2.5. Comparison between CBS and SICER. (A) Histograms of LAD coverage (coverage of Sicer based LADs by CBS based LADs (left) and vice versa (right)). Pink bars highlight LADs with zero coverage. (B) DamID traces showing visual comparisons of the LAD segmentation algorithms Sicer and CBS at the Igh LAD. Again, the raw data is displayed in the form of a histogram on a $\log_2$ scale in which a darker shade represents a positive $\log_2$ score for LmnB1/Dam. Previously published reads were mapped using our analysis pipeline (olive) and segmented using either Sicer (olive and above raw data) or CBS (blue, below raw data). Our MEF-seq data and identified Igh LAD is displayed in maroon. (C) Plots of the transition of average lamin B1 occupancy across all LAD borders (white: outside LADs, gray: inside LADs).

or epigenetic/nuclear structure data (ChIP-seq, ChiaPET, HiC, Bisulfite-seq). The internal truncations of theses LADs is illustrated using the LAD spanning the Immunoglobulin heavy (Igh) chain locus (Figure 2.5 B). While both the NKI MEFs LADs (array) and our MEF LADs (seq) pipelines identified interactions that span the

entire Igh locus, the SICER-mediated identification of sLADs shows this region to be comprised of many truncated "LADlets", in spite of robust signal across the locus. This can also be illustrated by comparing the SICER-mediated sLADs with their respective raw data genome-wide. Once again, we plotted genome wide averages of lamin B1 occupancy extending from -2.5kb to 2.5kb of s/LAD borders with the aid of the R based, Genomation package (Figure 2.5C). There is no clear LAD on either side of the boundaries, even though these data are from their own identified LADs. In additon, the 'peak' rather than 'domain' type of detection is evident. Unlike HMM-mediated and CBS-mediated LAD identification, which do identify domains and correlate with LaminB1 intensities. The LAD identifications provided by SICER are inconsistent with the LAD nomenclature, previous publications and internal measures and SICER is therefore not suited for LAD segmentation.

We next took this opportunity to benchmark another LAD segmentation program, Enriched Domain Detector (EDD, previously used for lamin ChIP) that has recently been published as an alternative to SICER in delimiting LAD domains (Lund et al., 2014). Since there is no published MEF LAD data using EDD, we used our data to test the algorithm. The Enriched Domain Detector was run with recommended settings and allowing the algorithm to determine the best bin size. Using the LADs identified by EDD or our CBS-based algorithm and their respective raw data, we plotted genome wide averages of lamin B1 occupancy extending from -2.5kb to 2.5kb of LAD borders with the aid of the R based, Genomation package (Figure 2.6A). While EDD is far superior when compared to SICER, it does not display the steep transition of lamin B1 occupancy at LAD borders as compared to HMM and CBS-based algorithms (Figure 2.6A, 2.5C,

Figure 2.6. Comparison between CBS and EDD based LAD segmentation algorithms. (A) Histograms of LAD coverage (coverage of EDD based LADs by CBS based LADs (left) and vice versa (right)). Pink bars highlight LADs with zero coverage. (B) DamID traces showing visual comparisons of the LAD segmentation algorithms EDD and CBS. The raw data is displayed in the form of a histogram on a $\log_2$ scale in which a darker shade represents a positive $\log_2$ score for LmnB1/Dam. We segmented our data using EDD (blue) and CBS (maroon). (C) Plots of the transition of average lamin B1 occupancy across all LAD borders (white: outside LADs, gray: inside LADs).



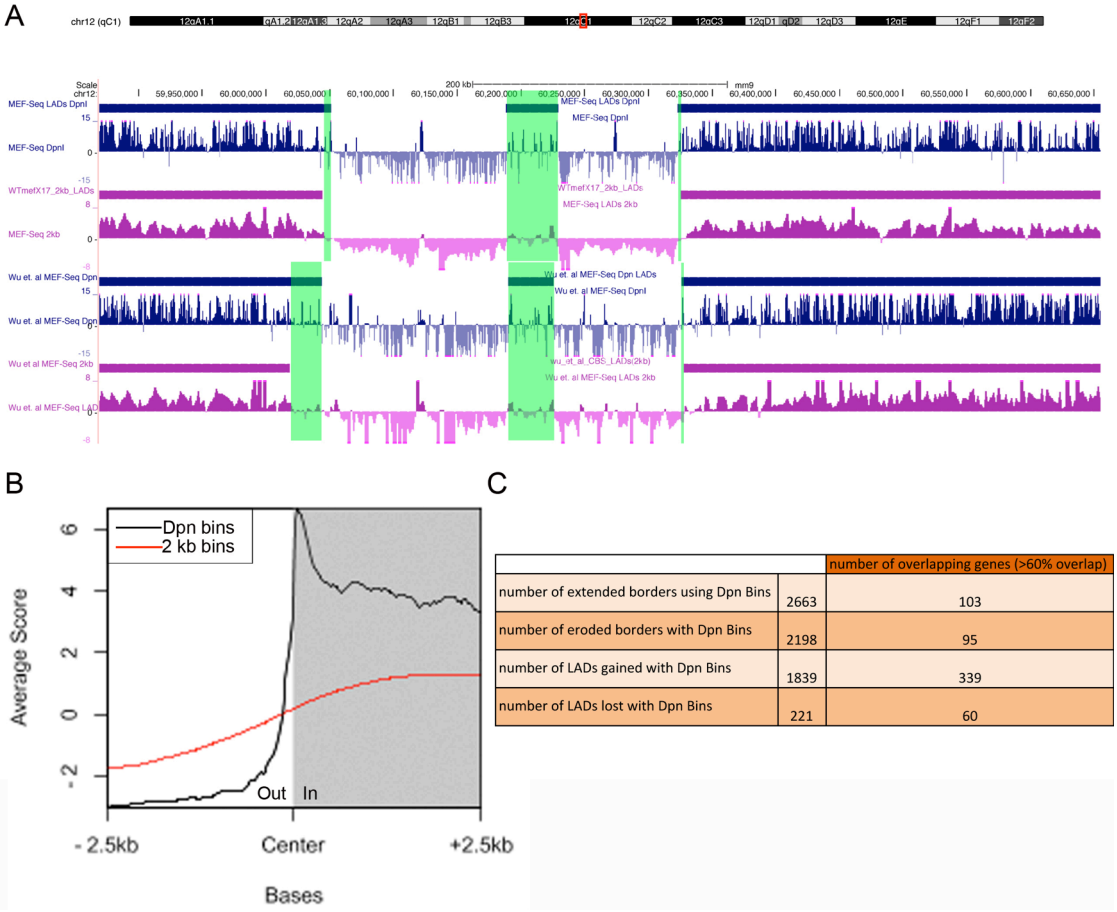
2.4B). Close inspection of the EDD LADs and the raw data reveals that the EDD determined LADs very often truncate before the end of a LAD is reached (Figure 2.6B). Additionally, some LADs are not detected by EDD, although not as extensive a loss as SICER (Figure 2.6C). .We find that both the Circular Binary Segmentation and Hidden

Markov Model based algorithm show the steepest transition in lamin B1 occupancy across LAD borders. Additional tweaks of the EDD parameters did not improve coverage, without losing boundary information (data not shown, Lund et al., 2014).

Based on these data, we conclude that segmentation of LADs via our CBS-based algorithm is quite robust. In summary, both CBS or HMM based algorithms are superior for accurately detecting LAD boundaries compared to either SICER or EDD. We have made our CBS-based algorithm available as a script for download.

In addition to various algorithms used to identify LADs, several groups have binned the genome in predetermined increments, of usually 2 Kb. Because we wanted to reflect the actual resolution of the technique (based upon GATC sites) in our non-LAD/LAD border identification, we next compared the use of *Dpn*I (or GATC) bins with contiguous 2kb segments on our DamID-seq data,. To do this, we assigned scores for each 2kb window in a manner similar to that described for our DpnI (Figure 2.2) bins and compared the resulting two traces visually, UCSC genome browser view, and bioinformatically (Fig 2.7). Although the chromosome-wide profiles look similar between the traces binned by GATCs and those binned by 2kb windows, the dynamic range diminished when we utilize 2kb contiguous windows as opposed to *Dpn*I bins. Notably, the LAD boundaries become less sharp with the use of 2kb windows and as a consequence, affected LAD segmentation, particularly at LAD borders – causing some LADs to be prematurely truncated while other LADs become extended . Additionally, some smaller LADs are lost while others are gained as a result of the smoothing effect when using 2kb bins. This observation holds when we perform the same analyses on an independent MEF DamID-seq dataset, in this case, reads from GSE41583 from the Yao

lab. We applied our bioinformatics workflow to these data (Figure 2.7). Combined, these data indicate that using 2Kb bins, as previously described in the literature, leads to erosion or extension of identified LAD boundaries. This has important ramifications in downstream analyses, as



		number of overlapping genes (>60% overlap)
number of extended borders using Dpn Bins	2663	103
number of eroded borders with Dpn Bins	2198	95
number of LADs gained with Dpn Bins	1839	339
number of LADs lost with Dpn Bins	221	60

Figure 2.7. Smoothing Dam-ID data affects the precision of LAD segmentation at LAD borders. (A) Shown in blue are Dam-ID sequencing data utilizing Dpn bins, with the segmented LADs above the $\log_2$ ratios. Shown in pink are Dam-ID sequencing data utilizing 2kb contiguous bins, with the segmented LADs above the smoothed $\log_2$ ratios. Our data is shown on the top and the bottom shows processed data from raw reads taken from GSE41583. The use of 2kb bins cause some LAD borders to become extended or "eroded" highlighted in green or LADs gained or lost, again highlighted in green. (B) shows a plot of average lamin B1 occupancy (in black, averaged from data using Dpn bins and in red, averaged from data using 2kb contiguous bins) across all LAD borders. (B) shows that the use of 2kb bins drastically changed the steepness of transition at LAD borders, hence affecting the LAD segmentation by circular binary segmentation. (C) shows the number of genes that overlap with the loss/gain of LADs or the extension/erosion of LADs as a consequence of data smoothing.

LAD boundaries have been implicated in organization of these domains and to harbor the majority of variable LADs. Even though LADs have been documented to be gene poor, we have previously documented that the borders of LADs tend to be highly enriched in developmentally regulated genes. Consequently, it is of paramount importance that the borders of LADs be as precise as possible during segmentation to study the dynamics of developmental gene regulation during differentiation and their perturbations in a disease state.

2.4 Discerning lamina and non-lamina association of repetitive elements

It has been documented that LADs can be subclassed into cLADs (constitutive/common) and f/vLADs (facultative/variable) (Luperchio et al., 2014; Meuleman et al., 2013). This sort of classification is inspired by the observation that local changes occur during cellular differentiation (vLADs), with respect to the formation of new LADs and the release of certain LADs/LAD domains from the periphery, on top of a seemingly less dynamic core genome architecture (cLADs). Additionally, genes that lie in regions that become a new LAD become silenced while genes in LADs/LAD domains that become released from the periphery become activated or poised for activation highlighting the importance of the nuclear periphery in regulating developmentally specific gene expression programs.

In characterizing the differences between cLADs, vLADs and non-LADs, we and others have documented a relative enrichment and depletion of LINEs and SINEs, respectively, within LADs (Meuleman et al., 2013; Zullo et al., 2012). However, most of this work derives from somewhat correlative measures that contain implicit assumptions.

In particular, previous documentations of repeat element enrichment or depletion are based on tiled array data which have been masked for these very repeats. As such, a measure of percent coverage of LADs by a certain repeat element implicitly assumes that repeat masked regions of the genome have attributes similar to regions immediately flanking it, i.e. if a repeat masked region lies between adjacent regions identified as a LAD, that region is by default also determined to be lamina proximal. In addition, no study to date has determined the relative enrichment of centromeric or peri-centromeric repeats in relation to LADs.

In this study, we determined the association of repeat elements with the nuclear periphery using two independent approaches. In the first approach the coordinates of each repeat type are downloaded from the UCSC table browser as a .bed file. Each repeat type is associated with multiple instances that correspond to multiple locations in the genome. We used a coordinate intersection approach to obtain a score of the percent coverage of LADs for a particular repeat element (all instances) and normalized this by the percent coverage of non-LADs for the same repeat. This approach is similar to what has been previously documented, with the exception that our genome wide data, *Dpn*I bins, have not been repeat masked (Meuleman et al., 2013). In a second approach, we directly input quality trimmed reads of either the Dam or Dam-LaminB1 samples into the RepeatMasker (RM) software (http://www.repeatmasker.org/) to discern the number of occurrences of specific repeat elements in these libraries. The number of occurrences of each repeat is then normalized by the read counts of that sample. The normalized

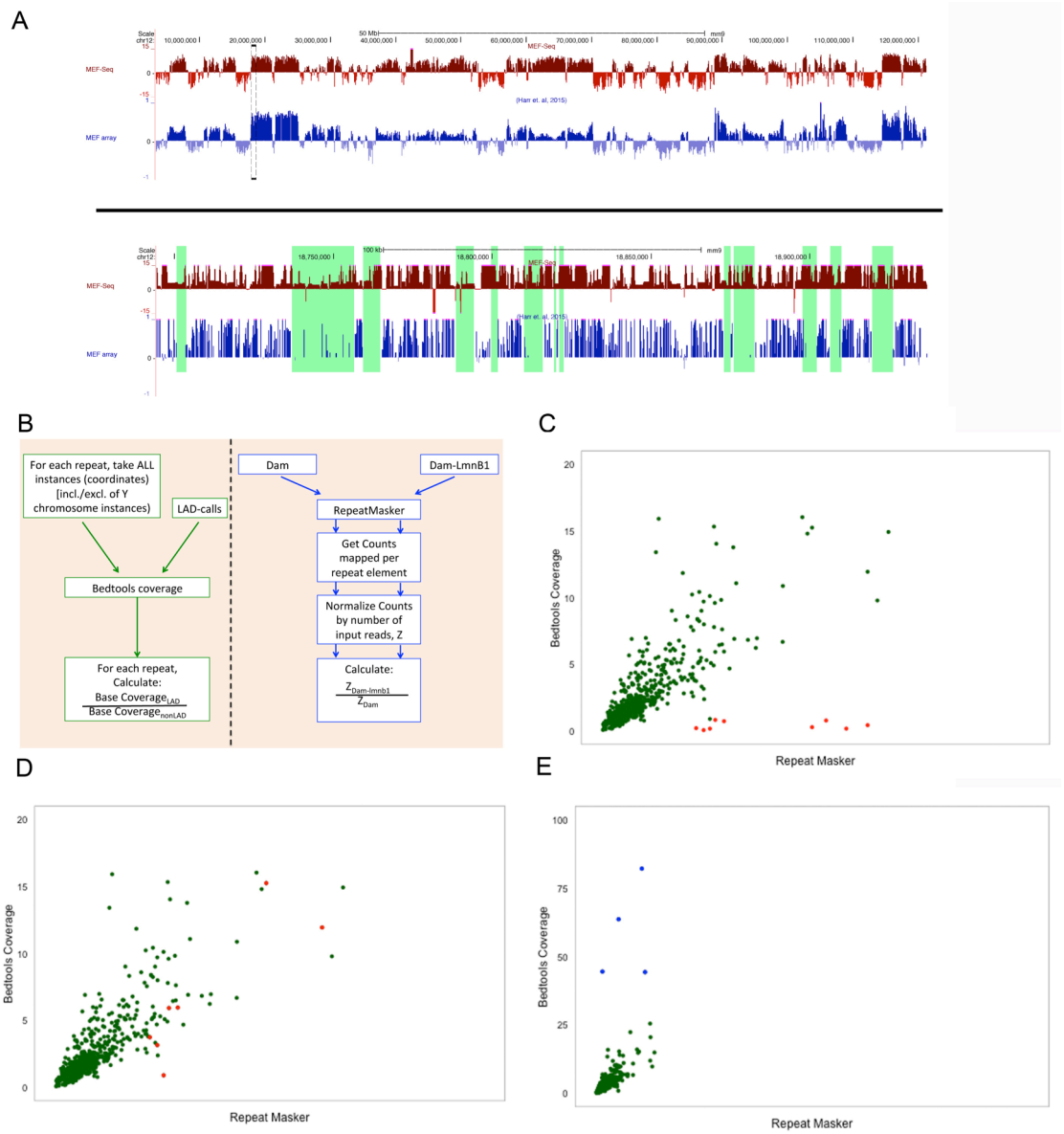


Figure 2.8. Comparison of sequence based (RepeatMasker) and coordinate based (Bedtools) repeat analysis. (A)Traces of DamID sequencing (Red), DamID array data of primary MEF cells (blue)-Top. Bottom - A zoomed in section of chromosome 12 shows large regions of the tiled arrays that are devoid of probes (highlighted in green) because tiled arrays are generally repeat masked to prevent non-specific cross reactivity of hybridization. (B) shows the parallel workflow for both analysis methods. (C) Each type of repeat element is plotted based on both the RepeatMasker and Bedtools score. Shown in red are obvious outliers. (D) All instances of each repeat element are filtered for coordinates that do not correspond to random chromosome contigs and re-analyzed using bedtools. Each type of repeat element is again plotted based on scores from RepeatMasker and the modified bedtools analysis. Shown in red are the previous outliers that are now showing high correlation of the 2 methods of analysis. (E) shows a zoomed out view of (D). Shown in blue are other outliers, including the GSAT_MM repeat element that persists and reasons for which are highlighted in the text.

score for the Dam-Lamin B1 sample is then divided by the normalized score for the Dam only control to yield an enrichment score. It is important to note that this approach is not reliant on mapping repeats or LADs to specific genomic coordinates, but instead directly measures the occurrence of repeats in the sequence data.

We notice that there is a reasonable correlation between the two methods of determining repeat enrichment or depletion within LADs. However, there are obvious differences between the two methods (Figure 2.8). Specifically, several repeat elements that yielded drastically different RM and Bedtools scores were selected for further investigation. Interestingly, these selected elements have many instances corresponding to "random" chromosome contigs which were not used to build the mm9 reference genome for bowtie mapping, upon which the bedtools analyses are reliant. Removal of instances overlapping to such contigs eliminated many of these extreme outliers except for MerSAT1, IAPLTR3-int and GSAT_MM. Interestingly, MerSAT1 and IAPLTR3-int have many instances overlapping the X chromosome. Because DamID, at least in this setting, is unable to discriminate between homologs, the X chromosome would appear in these genomic analyses to be heavily LAD-ed even though it has been widely documented that cells, of female origin, used in this particular analysis, consist of a highly heterochromatinized Barr body and another less compact X chromosome. As a consequence, the copious instances of MerSAT1 and IAPLTR3 that correspond to the X chromosome would skew these elements to become more enriched in LADs as compared to analyses utilizing RepeatMasker that implicitly takes into account sequences from both homologs. Lastly, GSAT_MM, gamma satellite sequences, occupy the pericentromeric region of chromosomes in immense copy numbers in a highly repetitive fashion. There

are, however, several instances of GSAT_MM that are not pericentromere related. Unfortunately, because the genome builds are poorly defined at the ends of chromosomes, the pericentromeric GSAT_MM coordinates are absent from the dataset downloaded from UCSC. The remaining instances map mostly to LADs hence yielding an appreciably high, but skewed, Bedtools score. Because the RepeatMasker takes into account all GSAT_MM sequences, which are heavily enriched at pericentromeres, a relatively low (close to 1) score was obtained, consistent with cell biology data showing chromocenters, which are clusters of pericentromeric heterochromatin, being interiorly positioned in the nucleus, with no bias to being at the nuclear periphery. Consistent with our previous data, these analyses showed that LINES are significantly enriched in LADs (Figure 2.9B). Intriguingly, our new approaches using DamID-seq and bedtools or RepeatMasker based approaches, show a significant enrichment of LTR type repeats and non-pericentromeric associated gamma-

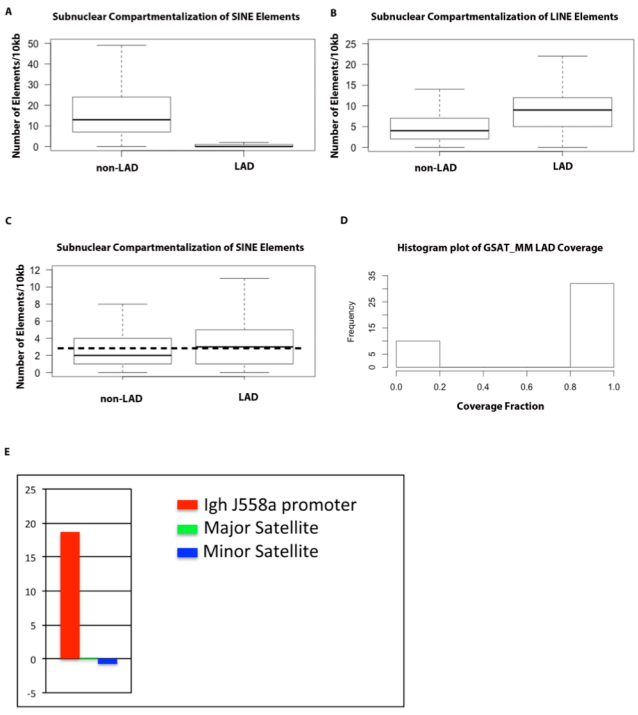


Figure 2.9. Subnuclear partitioning of several repeat classes. (A-C) shows the distribution for the density (number of elements per 10kb) of SINE, LINE and LTR elements in nonLADs and LADs. (D) shows the histogram of partitioning of non-pericentromeric GSAT_MM repeats. Coverage of >0.8: LAD, Coverage <0.2: nonLAD. (E) shows the $\log_2$ ratio of LaminB1/Dam for a region in the Igh locus, the major and minor satellite sequences.

satellite repeats (Figure 2.9 C, D), neither of which have been previously identified as enriched in these domains.

2.5 Discussion

In this chapter, we have described a method adapting DamID to a deep sequencing platform and have identified the best approaches for bioinformatically segmenting LAD domains from other genomic regions. In particular, we demonstrate that using Hidden Markov Model-based or Circular Binary Segmentation-related methods are superior at defining and delimiting LAD domains. Our CBS method worked very well in all benchmarking studies and proved to identify the most LADs, while also preserving robust LAD boundary identification. This algorithm is publicly available and can also be used to detect small breaks in LAD organization (Dips), that may be important for LAD regulation (data not shown). On the contrary, two methods used by other studies to identify LADs, namely SICER and Edge Domain Detection (EDD), failed to perform well in comparison. SICER -based analyses were particularly problematic, as they failed to detect domains at all, instead displaying numerous peaks across LADs. EDD mostly suffered from misrepresenting border regions and the loss of small LADs. Tweaking the program settings did little to alleviate these problems, as adjusting the parameters to identify these small LADs, led to over-extension of LADs and joining of adjacent clearly delimited LAD regions (data not shown and Lund et al., 2014).

Leveraging this sequence data has also revealed previously unappreciated types of repeat elements that are enriched in LADs. In particular, we show that in addition to the previously reported enrichment of LINE elements in LAD, LTR elements and non-pericentromeric gamma satellite repeats are preferentially in LADs. Finally, we

demonstrate that pericentromeric and centromeric repeats are not enriched in LAD domains, in agreement with cytological studies, but in contrast to previously published array-based DamID reports suggesting that LADs harbor such repeat elements (Meuleman et al., 2013). Since repeat elements are often associated with structural genome organization and are dysregulated in many diseases, it is important to understand how LAD organization relates to these elements and how dysregulation of LADs may impact these elements (and vice versa).

2.6 Materials and methods

Cell culture

C57BL/6-3T3 derived mouse fibroblasts were obtained from ATCC (CRL-1658) and were grown in DMEM + 10% fetal bovine serum at 37°C with 5% CO_2.

DamID library preparation

DamID was performed using the lentivirus transduction protocol as previously described (Guelen et al., 2008b). Briefly, mouse Lamin B1 cDNA was cloned into the pLGW-dam-V5 vectors kindly provided by the Van Steensel lab. HEK 293T cells were transfected with the DamID vectors and packaging plasmids (vSVG and Δ8.9). 12 hours post transfection, the transfection media was replaced with optiMEM. The media containing the lentiviral particles were collected every 12 hours for up to 48 hours. C57BL/6 cells were infected with the DamID lentiviruses that have been supplemented with polybrene at a concentration of 10ug/ml. The viral supernatant was removed after an overnight incubation and fresh media was added to the infected cells. These cells were then cultured for a minimum of 48 hours post infection.

Genomic DNA was harvested from cells expressing a Dam fusion protein and the Dam only control using the Qiaamp DNA minikit (Qiagen, 51304). For each sample or treatment, 2.5ug of DNA was digested overnight at 37 ^{o}C with DpnI (NEB, R0176S). DpnI was then heat inactivated at 80 ^{o}C for 20 minutes. This is followed by the ligation of DamID adapters (dsAdR) to the digested DNA, carried out at 16 ^{o}C for at least 4 hours. The ligase was then heat inactivated at 65 ^{o}C for 10 min. Subsequently, the ligated DNA was subjected to DpnII (NEB, R0543S) digest for 1 hour at 37 ^{o}C. DpnII digested DNA was then amplified via a ligation mediated PCR protocol with a primer that recognizes sequences interior of the adaptors, AdPCR. Post-DamID amplified material was randomized by performing an end repair reaction followed by a ligation and sonication. 1ug of column purified (QIAquick PCR Purification Kit: Qiagen, 28104) DamID material was end repaired using the NEBNext End Repair Module (NEB E6050S) following manufacturer's recommendations. Subsequently, the end-repaired DNA was purified, again, using the QIAquick PCR Purification Kit (Qiagen, 28104). 1ug of the resulting DNA was then ligated in a volume of 20uL with 1ul of T4 DNA ligase (Roche, 10799009001) at 16 ^{o}C overnight. Sonication was carried out using the Bioruptor® UCD-200 at high power, 30s on, 30s off for 1 hour in a 1.5ml microfuge tube (Sigma, T9661). The DNA was then transferred to 1.5 ml TPX tubes (Diagenode C30010010-1000) and sonicated for 4 rounds of 10 min long sonications with settings of high power, 30s on and 30s off. The DNA was transferred to new TPX tubes after each round to prevent etching of the TPX plastic. The sonication procedure should yield DNA sizes ranging from 100-300 bp. After sonication, the DNA was precipitated and

subjected to sequencing library preparation using the NEBNext Ultra DNA library prep kit for Illumina (NEB, E7370S).

Mapping of NGS reads

100 bp reads were mapped using an in-house written bioinformatics pipeline. First, each read was quality trimmed. This algorithm uses a sliding window of 3 nucleotides. The sliding window first advances from the 5' end of the read, removing the most 5' nucleotide until the mean of the scores of the nucleotides in a window is greater than 30. The sliding window then advances from the 3' end and does the same operation. The quality trimmed reads were then filtered for reads greater than 25 nucleotides in length. The output of the trimming software is then fed into another program that detects and removes AdPCR primer sequences introduced during the DamID randomization procedure. This program looks for oligonucleotide sequences corresponding to any DamID primer or adapter sequences. Typically, any read that contains the adapter sequences dsAdR are concatemers of the adaptors and are filtered away. Any read that has an internal AdPCR primer sequence is split into different reads with the sequence prior to the primer as one read and the one after the primer as another. The output of this program is then used in a series of mapping and post processing steps. Mapping onto the mouse genome assembly, mm9, was done using Bowtie1 with default settings. Importantly, during the randomization procedure, we might end up with truncated AdPCR sequences at either 5' or 3' ends that cannot readily be detected from the previous filtering programs. This results in a small portion of our, otherwise, mappable reads unused in the first bowtie. The unmapped reads from the first bowtie are then

processed by removing 13 nucleotides from the 5'end. These reads are then fed into Bowtie1 for the second time. Unmapped reads in the 2nd bowtie will then be used for 3' end processing and a final bowtie. However, we wish to recover the original sequence of the unmapped reads (ie non-5' trimmed sequences). Hence, the unmapped reads from the second bowtie were search for, in the unmapped reads of the first bowtie. Once these sequences have been retrieved, the 3' ends were trimmed by 13 nucleotides and passed through Bowtie1 for the last time. All 3 mapped sam files for each sample were concatenated and intersected using Bedtools with an interval file representing the mm9 genome that has been pre-binned by GATCs. The occurrences of the number of reads mapped to each bin is tabulated and normalized by the length of the bin and the number of reads for that sample. This score is then further normalized to the score of the respective bins obtained for the Dam only control to yield the relative occupancy of the protein of interest in each bin. A $\log_2$ ratio was obtained for each bin and stored in a bed file for display purposes and downstream bioinformatics analyses.

Determining Repeat Contents in Sequencing Reads using RepeatMasker

Quality filtered reads (process of quality filtering described in the section Mapping of NGS reads) were converted to fasta format and input into the RepeatMasker program using the following command: RepeatMasker –species mouse input.fa

Determining Repeat Contents in Sequencing Reads using Bedtools

Workflow for this approach is shown in Figure 2.8.

LAD Segmentation

A modified circular binary segmentation using the DNAcopy R package (Seshan and Olshen, 2014), which is an algorithm for identifying copy number difference, which "tests for change-points using a maximal t-statistic with a permutation reference distribution to obtain the corresponding P-value" (Venkatraman and Olshen, 2007). A sliding window approach with a window size of 2 kb was used to combine neighboring segments, using in-house perl scripts.

Chapter 3

Role of A type Lamins in Genome Organization

3.1 Introduction

The nucleus is a highly organized and complex organelle responsible for housing the genetic content of the cell and coordinating the precise spatial and temporal regulation of transcription, among other important activities, during development and normal cellular functions. Recent evidence suggests that gene regulation can be influenced by nuclear architecture through large chromatin domain establishment as well as local enrichment of regulatory and structural factors (Cremer et al., 2001, 2006; Elcock and Bridger, 2010; Fedorova and Zink, 2008; Ferrai et al., 2010; Misteli, 2005; Scaffidi and Misteli, 2006; Van Bortle and Corces, 2012). In particular, the nuclear periphery has been implicated in chromosome organization and gene regulation. This domain is comprised of the dual membrane nuclear envelope (NE) that is studded with Nuclear Pore Complexes (NPC), an underlying lamina meshwork and a diverse repertoire of inner nuclear membrane (INM) proteins.– While many of the proteins at the nuclear periphery perform diverse functions, several of them such as LBR, Emerin and Lap2, have been shown to interact with transcriptional repressors, chromatin modifiers and core histones (Guarda et al., 2009; Haraguchi et al., 2004; Holaska et al., 2003; Ye, 1997; Zullo et al., 2012). .Electron micrographs and even light microscopy studies have demonstrated that a substantial amount of heterochromatin abuts this region, suggesting that this domain may play a role in establishing or maintaining heterochromatin formation and thus transcriptional silencing (Boveri, 1914; Comings, 1980; Cremer and Cremer, 2009, 2006; Heitz, 1929). Remarkably, studies have shown that the targeting of ectopic genomic loci, via forced tethering or through LASes (Lamina Associated Sequences), to the nuclear lamina drastically reduces the transcriptional output of that locus, in agreement with the

hypothesis that the INM/lamina serves as a site of transcriptional repression (Finlan et al., 2008; Jennifer C Harr et al., 2015; Reddy et al., 2008; Zullo et al., 2012).

Recent molecular mapping of chromatin contacts at the nuclear periphery by DNA Adenine Methyltransferase (DamID, see chapter 2) has identified large chromatin domains termed lamina associated domains (LADs: 0.1-10Mb) that display important cell type specific dynamics at the nuclear periphery. Specifically, genes within domains that gain association with the nuclear lamina mostly become repressed and genes within domains that lost peripheral contact become activated or poised for expression, corroborating cytological studies correlating the proximity of specific gene loci relative to the nuclear lamina to their transcriptional repression in several developmental systems (Kosak et al., 2002; Meister et al., 2010; Peric-Hupkes et al., 2010; Szczerbal et al., 2009; Williams et al., 2006; Yao et al., 2011). However, to date, very little data has been obtained on the role that lamins and INM proteins, the components of the nuclear periphery, play in LAD organization and function.

Interestingly, mutations in proteins of the nuclear periphery, in particular LMNA, result in developmental diseases that cover a spectrum of phenotypes ranging from tissue anomalies such as muscular dystrophies and lipodystrophies to premature aging (Bonne et al., 1999; De Sandre-Giovannoli et al., 2002; Eriksson et al., 2003; Fatkin et al., 1999; Holt et al., 2001; Lin and Worman, 1993; Magracheva et al., 2009; Muchir et al., 2000; Novelli et al., 2002; Raffaele Di Barletta et al., 2000). These diseases have been collectively termed the laminopathies. At this point, it is unknown how these proteins, and in particular the lamins, regulate genome organization. Moreover, it is unknown if genome dis-organization plays a role in any of the over 500 diseases associated with

mutations in proteins found in this region. Hitherto, most, if not all, studies on the roles LMN A/C play in normal developmental progression, in normal cell function, mostly investigate Lmn A and Lmn C collectively, with the underlying assumption that these two spliceoforms are functionally redundant (Al-Saaidi and Bross, 2015). However, our recent work investigating the mechanism of de novo LAD formation showed that cells depleted of both Lmn A and Lmn C, but not cells removed only of Lmn A, showed perturbed targeting to and maintenance of genic regions at the nuclear lamina (J. C. Harr et al., 2015).

These data raise the question of whether LmnC is uniquely important for anchoring of LADs at the periphery. To determine if this was the case, or if the difference we observed was an effect of haploinsufficiency due to the differential depletion levels of total LmnA/C protein, we developed shRNA mediated knockdown constructs specific for LmnA or LmnC. Unlike cells derived from LmnA only or LmnC only mouse models, in which either LmnA or LmnC has been removed by germline deletion, the shRNA constructs allow us to test the immediate consequences of obliterating either LmnA or LmnC (Coffinier et al., 2010; Fong et al., 2006). In our investigations, we found that LmnC, not LmnA, is important in instructing genome organization. Specifically, using the Tagged Chromosomal Insertion System (TCIS), in which a specific genomic site has been targeted to the nuclear periphery, we found that knocking down LmnC but not LmnA perturbed locus positioning at the nuclear lamina. Similarly, we show that knocking down LmnC and not LmnA perturbs overall chromosome organization using three dimensional Immunoflurescent *In Situ* Hybridization (3D ImmunoFISH) with specially designed chromosome conformation paints. Because LmnC might be exerting

its effect in interphase and/or through the mitosis we further tested the relationship of LmnC to chromatin throughout the cell cycle, relative to LmnA and LmnB1.. We find that LmnC lags behind both LmnA and B1 in its association with the periphery and remains in large puncta in the nucleoplasm into early G1. This coupled with the spatiotemporal correlation of post mitotic re-organization of late replicating domains, including LADs, and evidence that genome organization is plastic in early G1, suggest that LmnC may be involved in the re-establishment of LADs post-mitosis (Chubb et al., 2002; Dileep et al., 2015; Dimitrova and Gilbert, 1999; Walter et al., 2003). Finally, we performed interactome studies of lamins A and C (protocol and technical details discussed in chapter 3) and observed that LmnC binds more preferentially to chromatin remodeling proteins and several mitotic chromatin remodeling partners. Taken together, our results support the idea that LmnC helps to organize chromatin. Importantly, these studies represent the first functional study highlighting differences between LmnA and LmnC in genome organization, raising the possibility that these two spliceoforms may be differentially affected in the various laminopathies and should thus be studied separately.

3.2 Results

3.2.1 LmnC is specifically important in the peripheral position of Lamina Associated Sequences (LASes)

The expression of A type lamins, unlike the B type, is correlated temporally with developmental progression, being minimal or near undetectable levels in embryonic stem cells to being one of the more abundant nuclear proteins in most terminally differentiated cells (Constantinescu et al., 2006; Eckersley-Maslin et al.; Solovei et al., 2013). As such, the A type lamins have been thought to be instructive to development, with chromatin

organization being one of many possible modes of action. More recently, the role that LmnA/C plays in genome organization has become more evident. In a recent study, both Lamin B Receptor (LBR) and LmnA/C were shown to be necessary molecular tethers for heterochromatin at the nuclear periphery, with a role for LBR in early development (when LmnA/C levels are low) and LmnA/C taking on a predominant role later in development or in tissues/cells with low LBR levels (Solovei et al., 2013).

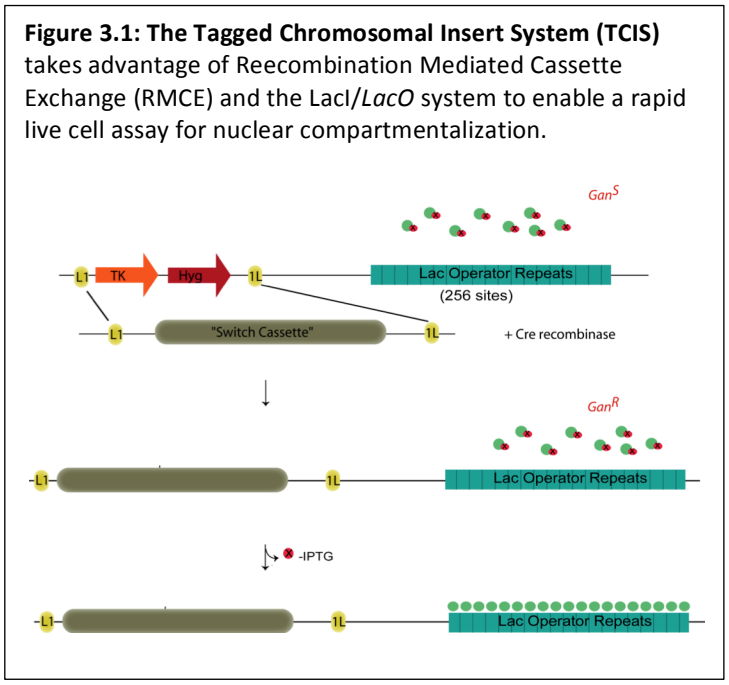


Figure 3.1: The Tagged Chromosomal Insert System (TCIS) takes advantage of Reecombination Mediated Cassette Exchange (RMCE) and the LacI/*LacO* system to enable a rapid live cell assay for nuclear compartmentalization.

To test the roles of LmnA and C in genome organization, we utilized the TCIS system described in Figure 3.1. (J. C. Harr et al., 2015). The Tagged Chromosomal Insertion Site (TCIS) system enables us to test the ability of discrete sequences to direct compartmentalization. Specifically, the TCIS system incorporates the ability to tag a chromosomal region via the *lacO*/LacI system (Belmont et al., 1999) and to recombine specific sequences into the same location using Recombination Mediated Cassette Exchange (RMCE) (Feng et al., 1999). The engineered chromosomal insertion is comprised of 256 *lacO* sites with an adjacent pair of inverted *loxP* sites which support recombination of specific sequences into this site. The locus can be followed by introduction of GFP-LacI which binds to the iterated *lacO* sites and accumulates to

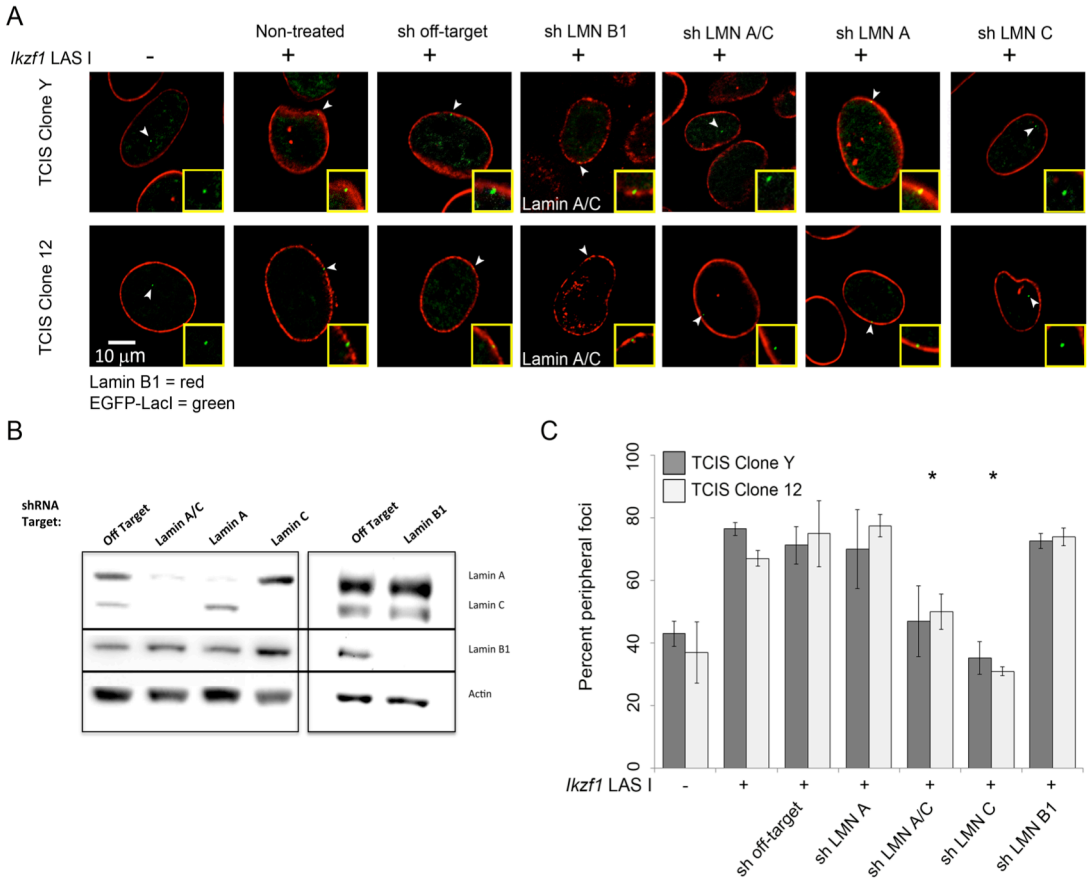
generate a discrete focus marking the insertion site. Specific sequences (such as sequences from LADs) in a "switch cassette" can then be introduced via Cre mediated recombination into this locus and the disposition of the chromosomal domain within the nucleus can be followed. As reported previously, we have two clonal TCIS lines, clones Y and 12, harboring a single copy of the TCIS cassette in a single genomic locus. Several LASes have been documented to be sufficient, when 'switched' into the TCIS locus, to cause the peripheral targeting of the TCIS cassette (J. C. Harr et al., 2015).

To investigate the effects of perturbing the levels of the different lamin proteins on lamina/peripheral association, we utilized versions of the TCIS clones that have been *a priori* positioned at the nuclear lamina, via the lamina associated sequence Ikzf1 D6 fragment (Ikzf1 Las1, Figure 3.2).

Consistent with previous findings, we demonstrate that depletion of LmnA/C perturbed peripheral association of lamina associated sequences, LASes (Figure 3.2, J. C. Harr et al., 2015). Interestingly, depletion LmnA levels alone has no effect on the peripheral disposition of LASes (Figure 3.2). These data led us to speculate that the LmnC isoform might be preferentially interacting with chromatin to help organize peripheral heterochromatin. However, we note that this difference may be due to either a specific need for LmnC at the lamina, or, might instead reflect that a loss of too much lamins, of whatever type, causing the loss of genome organization. For instance, there could be a minimal level of expression of A-type lamins (both LmnA and LmnC) required to maintain LASes at the periphery. This is supported by previous evidence that phenotypes of LmnA/C perturbations can be modulated with simply increasing LmnA/C levels (Sehgal et al., 2013). On the other hand, the LmnC isoform might be specifically

important in informing LAD establishment and/or maintenance at the periphery. To discriminate between these possibilities, we designed LmnA and LmnC specific shRNA constructs to independently perturb expression of these isoforms. The LmnA specific shRNA targeted a specific region of the coding mRNA that differs from LmnC enough to specifically perturb this isoform (Figure 3.2).

Figure 3.2: LmnC is necessary for peripheral positioning of LASes. (A) Representative images showing the disposition of lacO arrays (arrowheads, green) and LMNB1 (red) in the TCIS clones Y (top) and 12(bottom) pre- and post- "switching" in of the Ikzf LAS and the effects of specific lamin depletion on the position of the lacO arrays. The inset shows 300× magnification. (B) Representative western blots showing specificity of the various knockdown constructs. (C) Quantitation of peripheral association was determined by overlap of EGFP-LacI foci and LMNB1 (n ≥ 50). Error bars indicate SD. P ≤ 0.001 is indicated by asterisks.



Consistent with our previous findings, perturbing LmnA had no effect on the peripheral localization of the TCIS cassette whereas removing both A type lamins perturbed the maintenance of the TCIS cassette at the nuclear periphery, in both clones Y and 12. Interestingly, removing lamin B1, which has been routinely used (instead of LmnA or C) in discerning lamina associated domains, had no effect on the peripheral disposition of the TCIS cassette (Figure 3.2). Strikingly, knocking down LmnC alone perturbs localization of the TCIS cassette at the periphery, causing a substantial relocalization to the nuclear interior. This strongly suggests that the LmnC isoform, but not LmnA, is important for tethering LASes at the nuclear periphery.

To confirm that LmnC itself is sufficient to redirect the locus to the nuclear lamina, we next sought to rescue the peripheral association defects of LASes by overexpressing mCherry tagged human alleles of LmnA and LmnC. Consistent with the hypothesis that LmnC and not LmnA is important in the lamina juxtaposition of LASes, only mCherry tagged LmnC is able to re-inform the peripheral positioning of the TCIS cassette in both clones Y and 12 (Figure 3.3).

3.2.2 Genome Organization is affected by LmnC depletion

We next sought to investigate if perturbing the levels of LmnA, C or even total A-type lamins affected gross global nuclear architecture, particularly with respect to changes in the lamina associated domains (LADs). Initially, we used the DamID-seq technique to detect lamina-chromatin interactions (as described in Chapter 1).We rationalized that large perturbations in genome organization upon depletion of LmnC might result in obvious LAD landscape changes using this technique.

Surprisingly, we observed no change in global LAD architecture based upon bioinformatics analyses of LAD coverage and LAD border integrity by DamID (Figure 3.4). While this result was initially surprising, we rationalized that the DamID technique

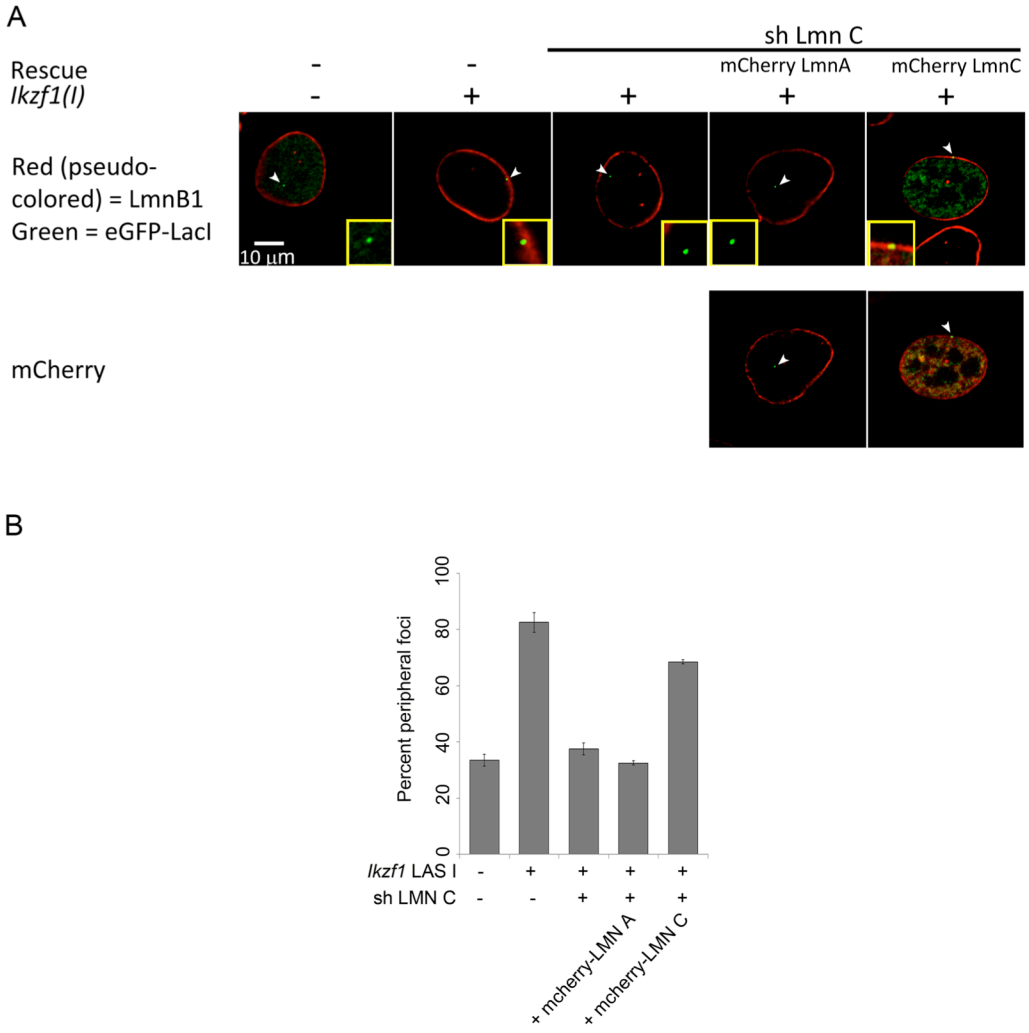


Figure 3.3: Re-expression of LmnC but not LmnA rescues the LAS tethering defect in LmnC deficient cells. (A) Representative images showing the disposition of lacO arrays (arrowheads, green) and LMNB1 (pseudo-colored red) in the TCIS clone Y. The inset shows 300× magnification. (B) Quantitation of peripheral association was determined by overlap of EGFP-LacI foci and LMNB1 ($n \geq 50$). Error bars indicate SD. $P \leq 0.001$ is indicated by asterisks.

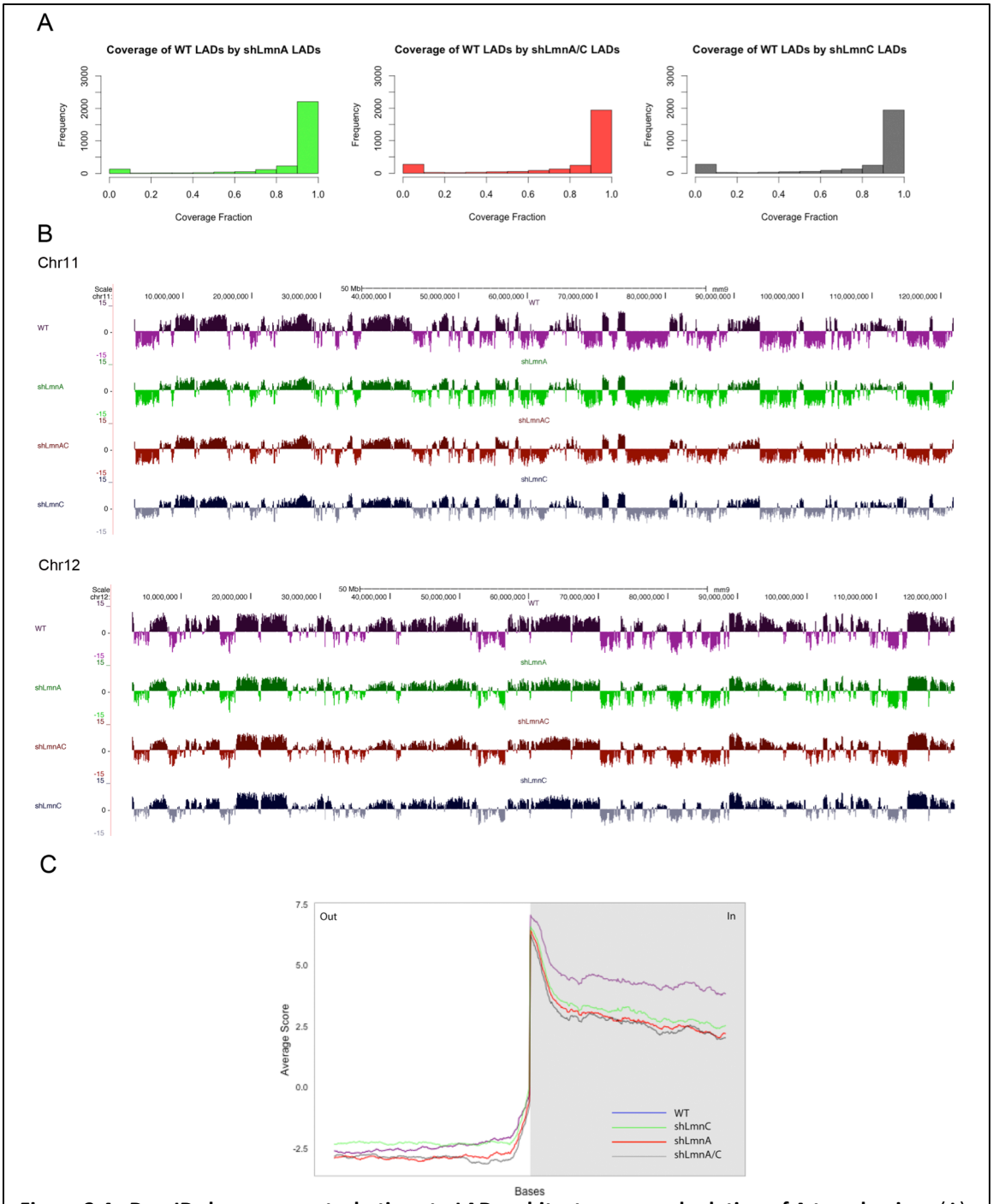


Figure 3.4. DamID shows no perturbations to LAD architecture upon depletion of A type lamins. (A) Histogram plots showing the coverage of LADs from untreated (WT) MEFs by LADs from samples perturbed for LmnA (green), LmnA/C (red) or LmnC (gray). (B) Chromosome wide DamID traces (chr 11 and 12) for wildtype (purple), LmnA depleted (green), LmnA/C depleted (red) and LmnC depleted (gray) samples. Vertical axes are of log2 scale and traces above 0, in darker shades, indicate a higher than expected frequency of peripheral association. (C) Profiles of aligned LAD border regions (left and mirrored right border regions combined) are shown for LMNB1 occupancy. To align LAD borders, genome-wide positions of LMNB1 occupancy were converted to coordinates relative to the nearest border. Gray area and positive coordinates, inside LADs; white area and negative coordinates, outside LADs.

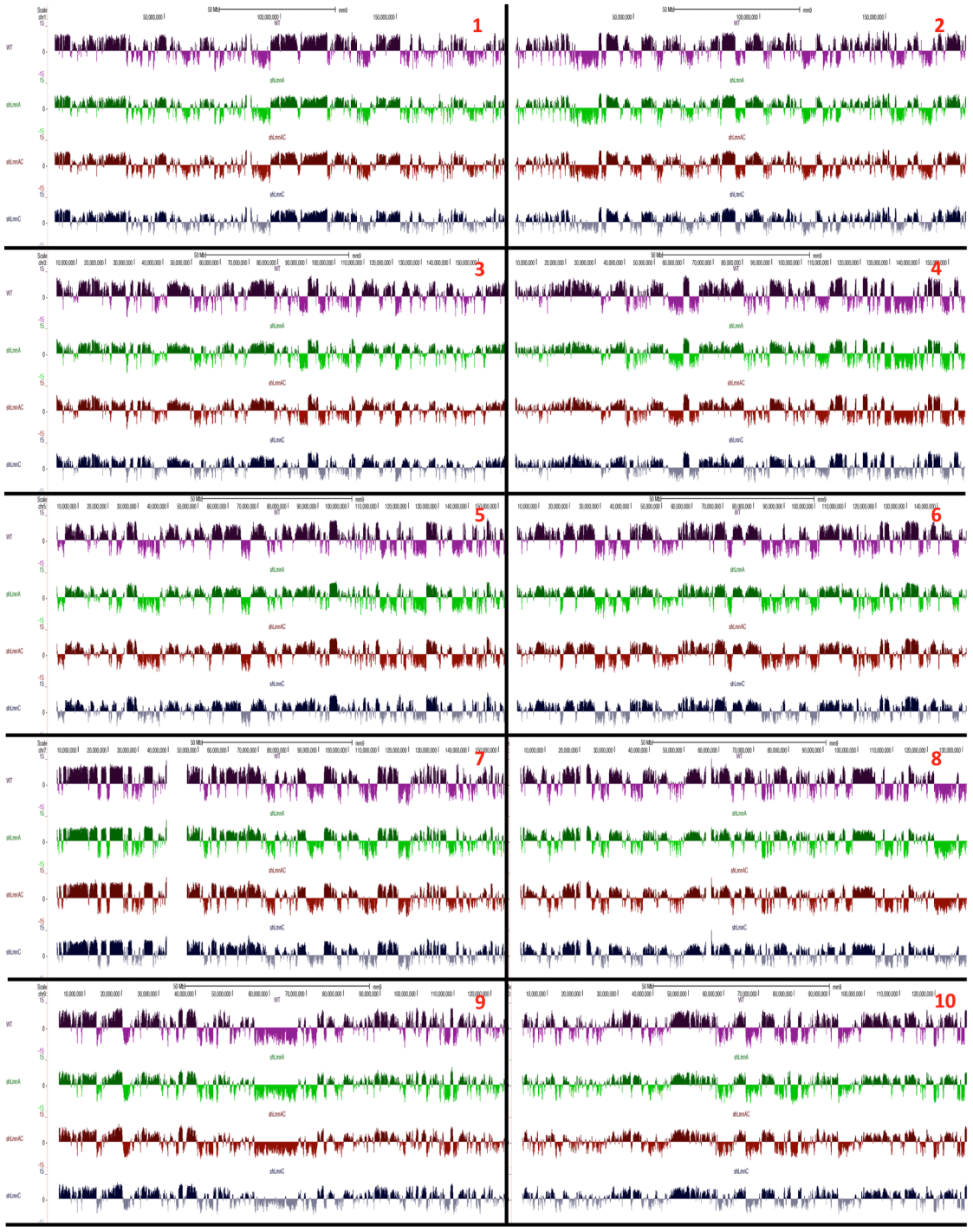


Figure 3.4 (continued). DamID shows no perturbations to LAD architecture upon depletion of A type lamins Chromosome wide DamID traces (all chromosomes) for wildtype (purple), LmnA depleted (green), LmnA/C depleted (red) and LmnC depleted (gray) samples. Vertical axes are of log2 scale and traces above 0, in darker shades, indicate a higher than expected frequency of peripheral association.

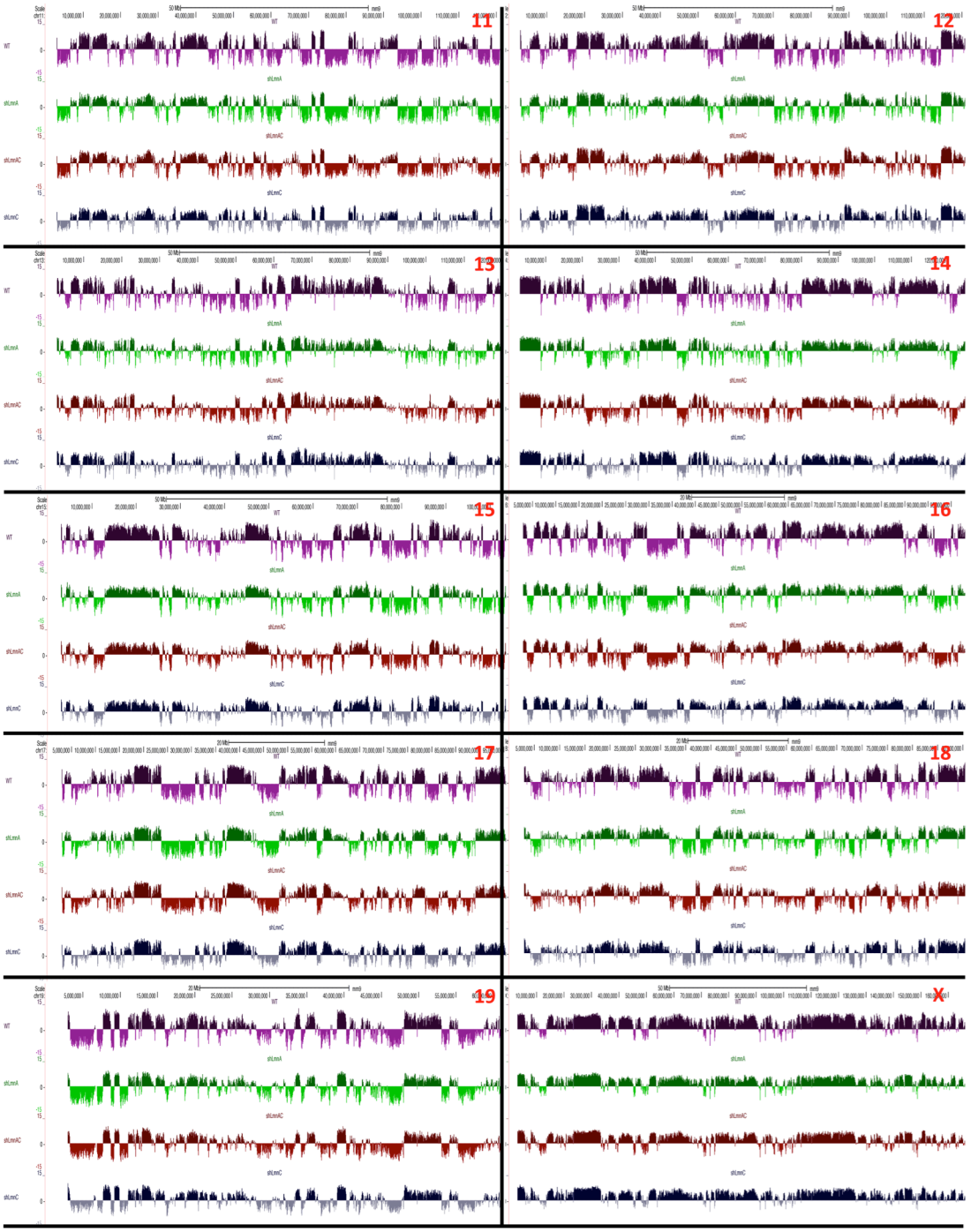


Figure 3.4 (continued). DamID shows no perturbations to LAD architecture upon depletion of A type lamins Chromosome wide DamID traces (all chromosomes) for wildtype (purple), LmnA depleted (green), LmnA/C depleted (red) and LmnC depleted (gray) samples. Vertical axes are of log2 scale and traces above 0, in darker shades, indicate a higher than expected frequency of peripheral association.

yields population signatures of positive associations with the protein of interest, in our case, the nuclear lamina protein lamin B1. Intermittent loss of interactions or a simple

decrease in levels of association may not show up robustly using this method. In addition, we note that shRNA mediated knockdown experiments are not fully penetrant, with some cells likely harboring significant levels of residual lamin targets. Additionally, perturbing a general mechanism of LAD establishment may not lead to all cells exhibiting synchronous and targeted disorganization. Thus, DamID may not be the best proxy to understand the genome wide, cell-to-cell, dynamic perturbations of lamin interactions in these populations.

Given the insensitivity of population-wide measures such as DamID or chromatin immunoprecipitation (ChIP) to cell-cell variation, we next utilized an approach that enables us to measure disruptions in chromosomal organization on a single cell basis. This analysis was performed using 3D Immuno-FISH with chromosome conformation probes (Figure 3.5). Briefly, population wide DamID data on wild-type murine embryonic fibroblasts (MEFs) was bioinformatically analyzed to segment LAD regions from the rest of the genome (non-LADs) (segmentation as described in Chapter 1). We designed chromosome paints by tiling approximately 150 nucleotide long oligonucleotides across the entirety of the repeat masked chromosome 11 and chromosome 12. These oligos were also designed to avoid sequences with high homology to off target loci and were additionally filtered for sequences with appropriate GC content and annealing temperatures. These probes were chemically synthesized by Agilent Technologies yielding approximately 288 million unique oligonucleotides for chromosome 11 and approximately 216 million unique oligonucleotides for chromosome 12. Prior to synthesis, the designed oligonucleotides for each chromosome were then separated into two pools depending on whether they corresponded to a genomic region

that was segmented as a LAD or a non-LAD. After synthesis, the two pools were then conjugated with different fluorophores, Cy3 for LADs and Cy5 for non-LADs. These two-color chromosome paints then reflect a bioinformatically derived conformation of chromosome organization. When these paints are applied to wild-type murine fibroblasts, fluorescent signals corresponding to LADs are positioned at the nuclear periphery, with the majority of LADs residing within 0.6um of the lamina (3. 5 and Luperchio et al. in preparation). Signals corresponding to the non-LADs however were further from the lamina. Importantly, signals from LADs and non-LADs segregated, showing for the first time, that chromosome configuration can be organized based on function. We therefore used this technique to investigate the effects of perturbing the A type lamins on functional chromosome organization at the single cell level. Preliminary data suggests that LmnA perturbations, consistent with data from single LASes, had no effect on chromosome organization. However, depleting LmnC levels had various effects on chromosome architecture. In some cases, chromosomes are seen to become detached from the nuclear periphery. In other instances, there appears to be intermingling of the LAD and non-LAD signals indicating a perturbation in functional chromosomal folding. These effects culminate in an overall dispersion of the LAD signal from the periphery as shown in Figure 3.5. Thus perturbation of LmnC specifically alters overall chromosome organization, with a loss of LAD organization at the lamina and, in extreme case, loss of LAD/non-LAD segregation.

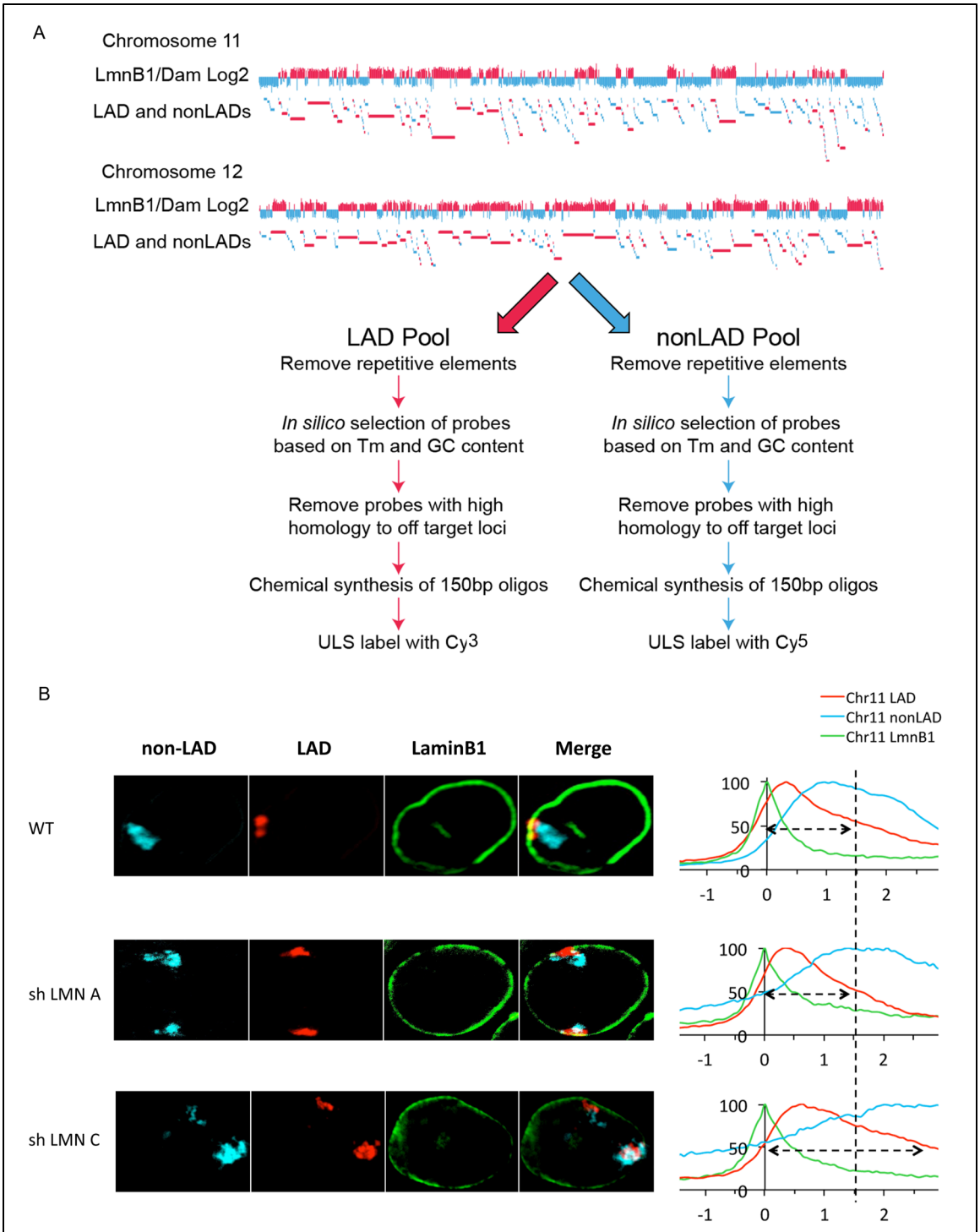


Figure 3.5. Nuclei depleted of LmnC but not LmnA show perturbations of chromosome organization. (A) T has the description for the probes. (B) Representative images showing the organization of chromosome 11 with nonLADs in cyan, LADs in red and lamin B1 in green (left). On the right are average normalized intensity plots for each condition (n=25 chromosome territories per condition). The average profile for a single territory was obtained by performing 3 lines scans, perpendicular to the lamina, across each territory. Shown in dotted arrow is the 50% max intensity mark for measuring the spread of the LAD signal with the different treatments.

In addition to probing for genome disruption, we also determined the effect of disruption of these lamin variants on gene expression. RNA was isolated from MEFS that had been subjected to shRNA-mediated knockdown of LmnA/C, LmnA or LmnC and processed for RNA-seq analyses (Heydarian et al., 2014). While the results of these experiments are still nascent, we find, consistent with our genome organization findings, perturbations of LmnA/C or LmnC alone caused more dysregulation than perturbing LmnA (data not shown). Importantly, using the publicly available Cluster 3.0 algorithm to determine how related two condition are to one another, we show that the LmnA knock down transcriptome clustered together with the wildtype transcriptome and interestingly, the transcriptome of LmnC depleted cells clustered with that of total Lmn A/C depleted cells. (Figure 3.6).

3.2.3 Temporal Correlation of LmnC peripheral recruitment and Genome reorganization post mitosis

We next investigated how LmnC may be informing genome architecture. In mammalian cells, which undergo open mitosis, interphase genome organization is apparently lost during mitosis to allow for chromosome condensation and proper segregation of daughter chromosomes. Post mitosis, therefore, the interphase genome architecture has to be re-established for proper cellular function. This large scale genome reconfiguration occurs in early G1, termed the 'timing decision point' (Chubb et al., 2002; Dileep et al., 2015; Dimitrova and Gilbert, 1999; Walter et al., 2003).. This brief period in early G1 has been speculated to be a window of opportunity for the integration of extracellular cues in informing cellular differentiation and establishment of cell type specific genome architecture (Gilbert, 2010). Consistently, several studies ectopically

recruiting nucleoplasmic loci to the lamina document the necessity for the passage through mitosis in appropriate lamina targeting (Finlan et al., 2008; Kumaran and Spector, 2008; Reddy et al., 2008). While studies comparing mitotic localization of various INM proteins as well as comparisons between LmnA and progerin, a mutant form of LmnA, have been documented, localization patterns for LmnC through mitosis have yet to be shown. We therefore performed a localization study investigating the differences LmnA and C might display in mitosis in order to gain insights to potential LmnC roles in post mitotic genome reconfiguration.

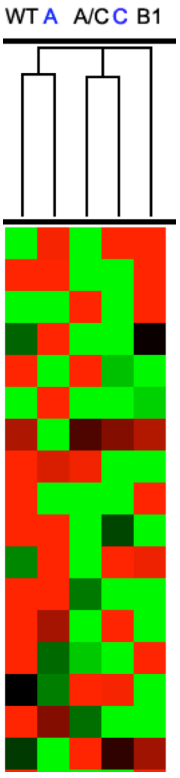


Figure 3.6 Hierarchical Clustering of transcriptome of various lamin type knock down cells. RNA-seq cluster data LmnC depleted cells are more disrupted than LmnA depleted cells. Shown in this figure is the dendogram based on FPKM values for expression of genes. The heatmap on the bottom shows the expression patterns of a selected set of genes

It has been widely documented that the nuclear lamins become phosphorylated at the onset of mitosis and their localization become diffused throughout the dividing cell. The nuclear envelope and lamina however, starts reforming at telophase as the chromosomes destined for the two daughter cells become fully segregated. Consistently, we have observed that the localization of LmnA, LmnC and LmnB1 become diffused during mitosis (Figure 3.7). Interestingly, during telophase, while LmnA and LmnB1 colocalize to form the nuclear lamina of the new daughter cells, LmnC displays a nucleoplasmic localization. This interior disposition persists through telophase and into part of G1 (Figure3.7).

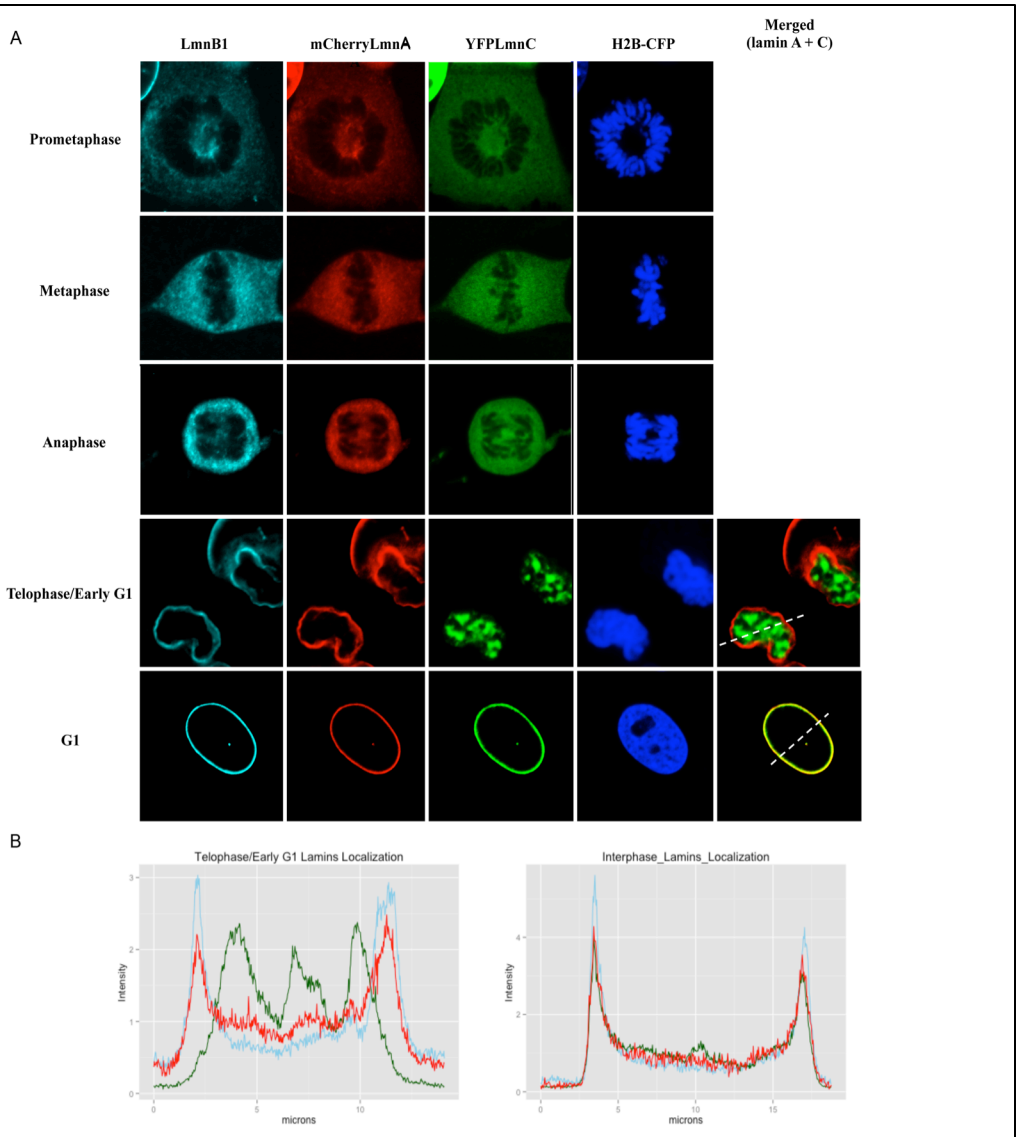


Figure 3.7. Differential localization of lamins A and C during mitotic exit. (A) Representative images of lamin B1(cyan), LmnA (red), LmnC (green) and chromatin (blue) in different stages of the cell cycle. Merged images for lamins A and C localization are shown in Telophase/Early G1 transition and G1 phases. Dotted lines on the merged images indicate the segment used for line scan displays shown in (B). (B) shows representative plots of the normalized intensities of lamin B1 (Cyan), LmnA (red) and LmnC (green) along the dotted lines (from left to right) as shown in the merged images in (A) for the telophase/early G1 transition stage of the cell cycle and the G1 stage.

Importantly, when we measure the association of LmnC versus LmnA with the nuclear lamina at the point of telophase/early G1, there is a striking and consistent difference. While in interphase, LmnA and LmnC both have strong peripheral localization (Figure

3.8), during early G1, LmnC lags behind on the reorganizing chromatin (Figure 3.9) .

3.3 Discussion

Taken together these data strongly suggest a role for LmnC in reorganization and targeting of Lamina Associated Domains and maintenance of genome organization. We find a particularly provocative role for LmnC in establishing genome organization post-mitosis. This is most intriguing because, coincident with the disintegration of the nuclear envelope and matrix, the dissolution of interphase genome architecture also occurs at the onset of mitosis, allowing replicated chromosomes to become compacted and segregated into subsequent daughter cells, as previously mentioned. Our data are in agreement with early cytogenetic studies reporting that the re-establishment of three dimensional repositioning of chromatin domains into their final interphase configuration, after mitotic exit, occurs at a specific time in interphase of the newly formed nuclei known as the timing decision point (Chubb et al., 2002; Dileep et al., 2015; Dimitrova and Gilbert, 1999; Walter et al., 2003. This observation has been supported by a recent study using the 4C-sequencing technique demonstrating that interphase chromatin organization is established in early G1, coincident with the timing decision point. (Dileep et al., 2015). Intriguingly, at this stage, some LADs are also still in the nuclear interior. We now speculate that the LmnC isoform plays a unique role in this reorganization. We note that the non-peripheral localization of LmnC, from telophase to early G1, is temporally conincident with the re-establishment of interphase organization suggesting that LmnC might play active roles in the re-establishment of genome architecture post mitosis

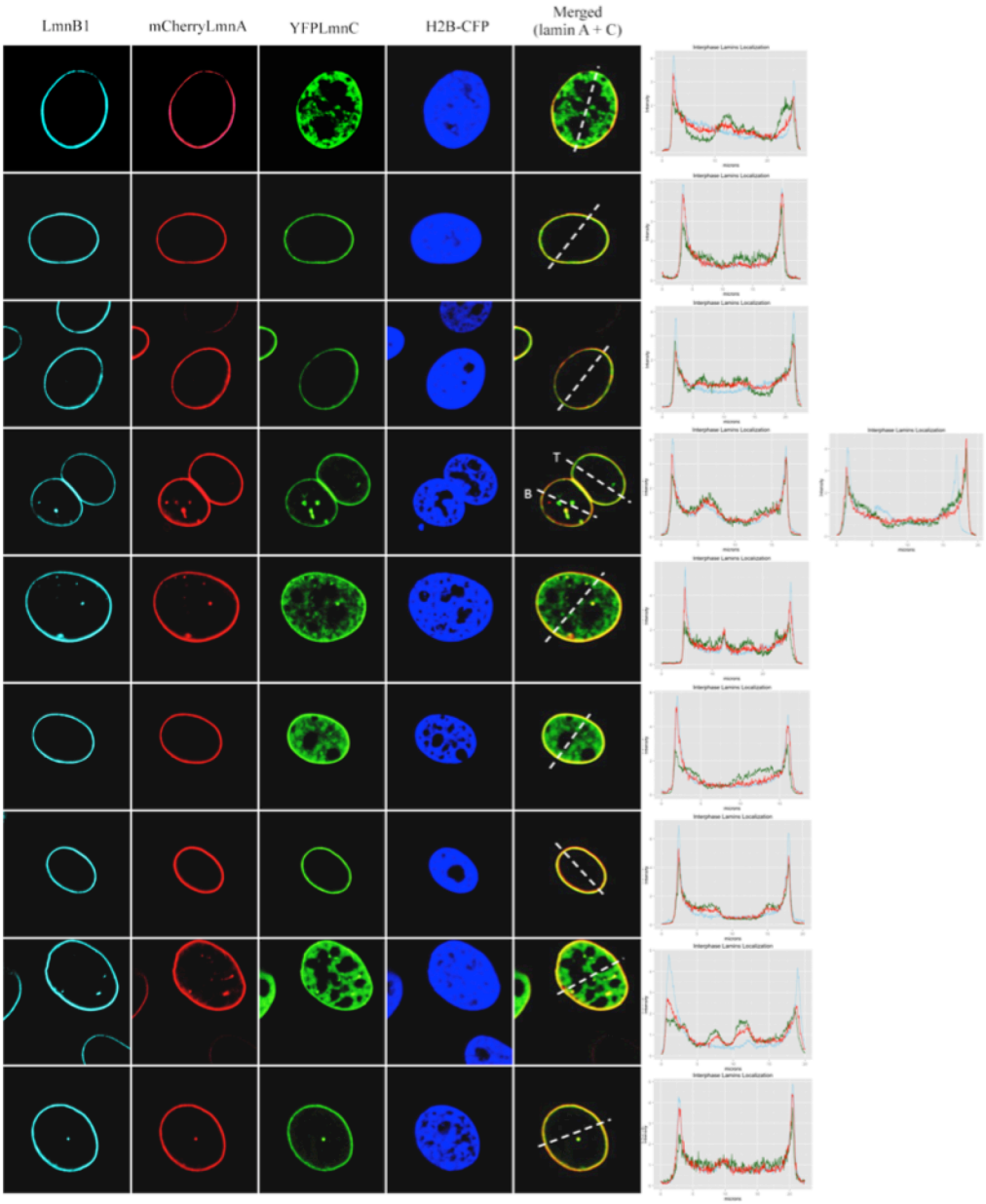


Figure 3.8 Localization of lamins during interphase: Images of lamin B1(cyan), LmnA (red), LmnC (green) and chromatin (blue) in interphase. Merged images for lamins A and C localization are also shown. Dotted lines on the merged images indicate the segment used for plotting the normalized intensities of lamin B1 (Cyan), LmnA (red) and LmnC (green) shown on the right.

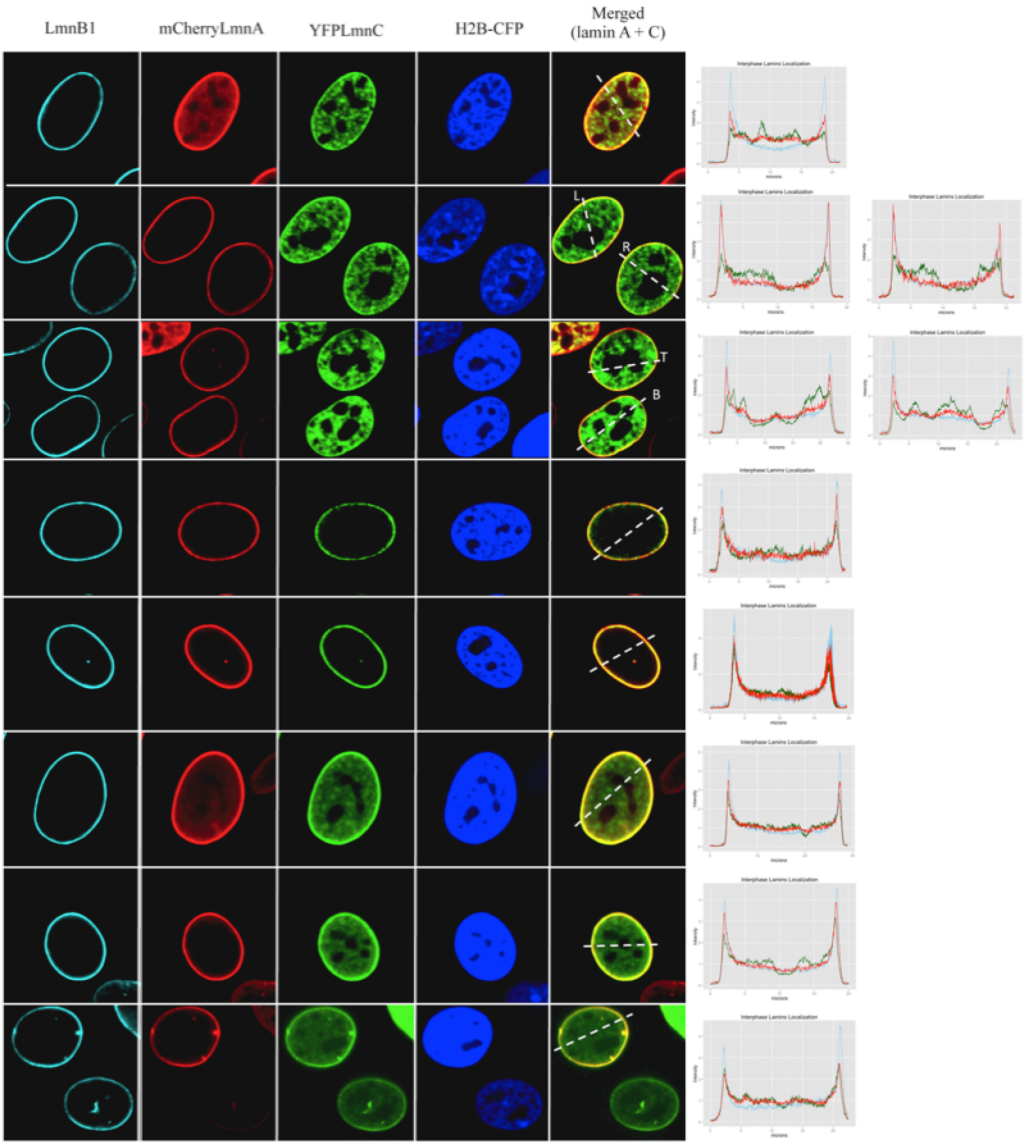


Figure 3.8 Localization of lamins during interphase (continued): Images of lamin B1(cyan), LmnA (red), LmnC (green) and chromatin (blue) in interphase. Merged images for lamins A and C localization are also shown. Dotted lines on the merged images indicate the segment used for plotting the normalized intensities of lamin B1 (Cyan), LmnA (red) and LmnC (green) shown on the right.

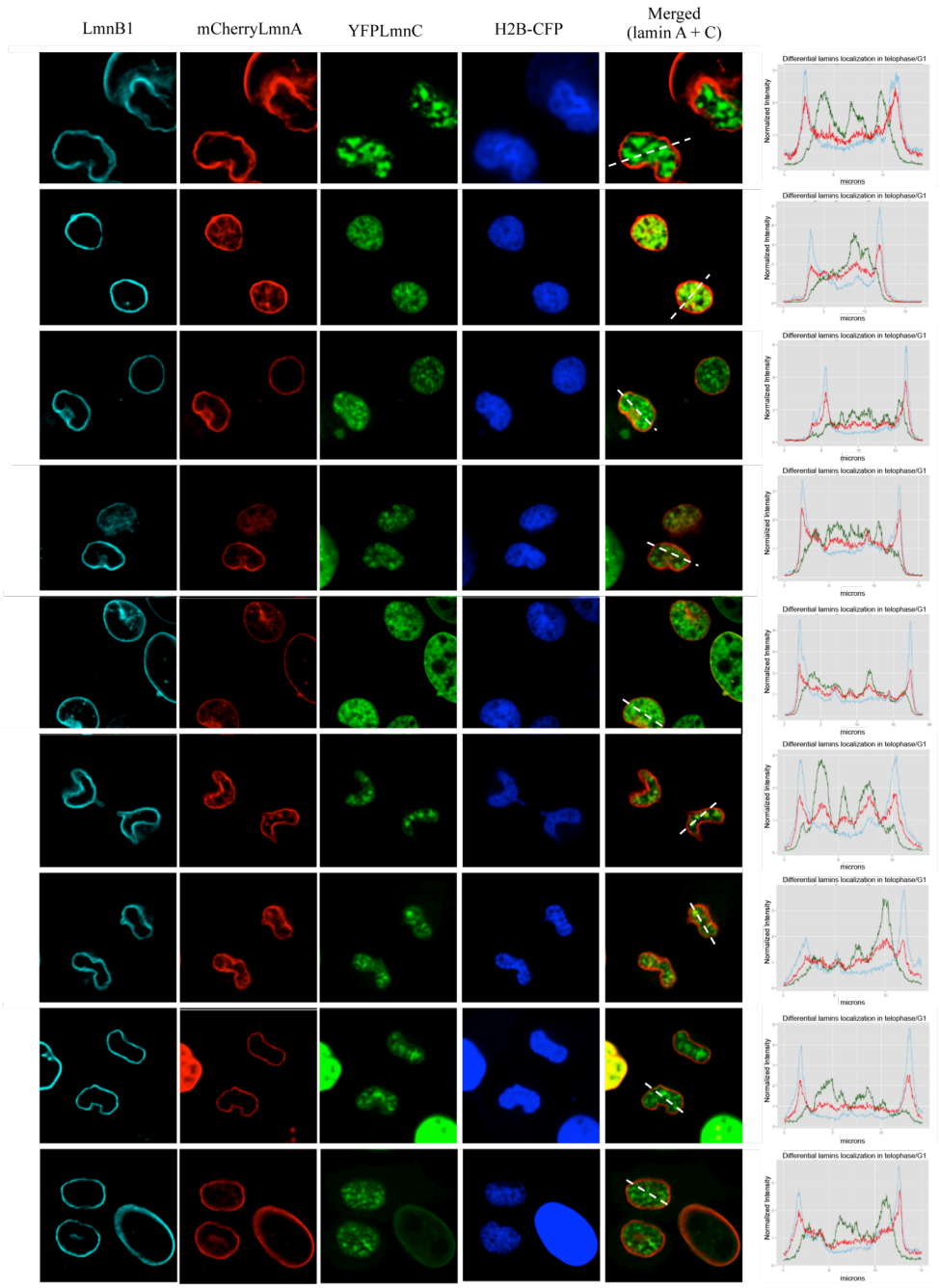


Figure 3.9 Localization of lamins during telophase/early G1. Images of lamin B1(cyan), LmnA (red), LmnC (green) and chromatin (blue) in telophase/G1 transition. Merged images for lamins A and C localization are also shown. Dotted lines on the merged images indicate the segment used for plotting the normalized intensities of lamin B1 (Cyan), LmnA (red) and LmnC (green) shown on the right.

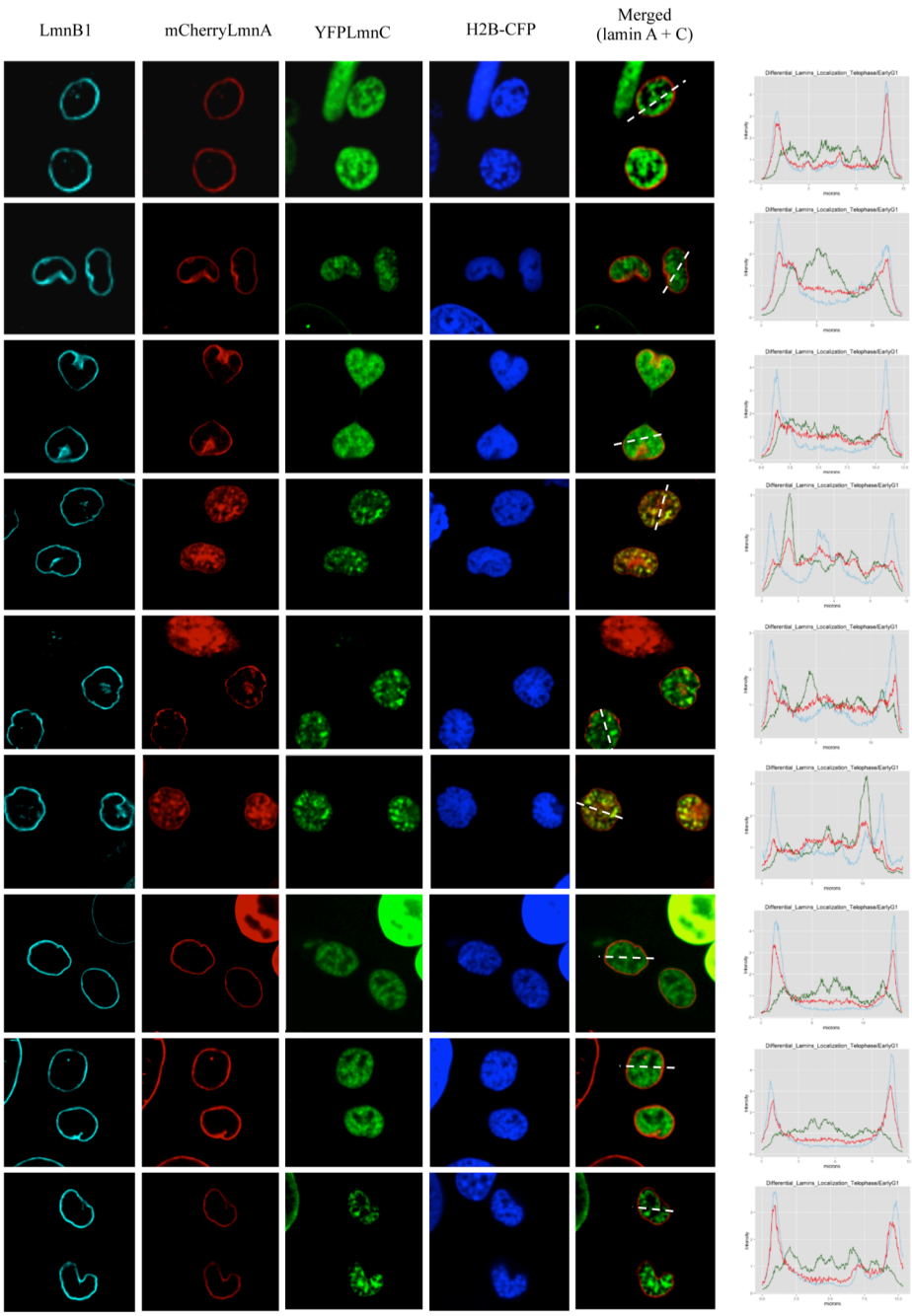


Figure 3.9 Localization of lamins during telophase/early G1 (continued): Images of lamin B1(cyan), LmnA (red), LmnC (green) and chromatin (blue) in telophase/G1 transition. Merged images for lamins A and C localization are also shown. Dotted lines on the merged images indicate the segment used for plotting the normalized intensities of lamin B1 (Cyan), LmnA (red) and LmnC (green) shown on the

Chapter Four

Identifying Differential Laminomes using Biotinylation Identification (BioID) coupled with SILAC Mass Spectrometry Technology

We next sought to understand the roles lamin C play in genome organization. While the literature appears to be replete with biochemical studies profiling lamin A interactors, data on lamin C interactions are not as comprehensive (Dittmer et al., 2014; Kubben et al.). Additionally, most lamin A interaction studies are either performed in vitro or rely on co-immunoprecipitation. While the former may not be indicative of in vivo interactions, the relative insolubility and high complexity of the nuclear lamina meshwork pose technical challenges both experimentally and analytically with respect to co-immunoprecipitation. Recently, the biotin ligase identification (BioID) method has been devised primarily to allow the identification of interacting partners in technically challenging biological compartments such as the nuclear lamina (Roux et al., 2012). We utilize this approach coupled with SILAC and mass spectrometry to identify differential interactors of lamin A and lamin C. Briefly, a mutant form of the Escherichia Coli biotin ligase, BirA*, is fused to lamin A and C. BirA* has been characterized to prematurely release labile activated biotin which then reacts with primary amines in the vicinity. This allows the in vivo labeling of protein partners of lamin A and C with biotin which can then be affinity purified in a relatively denaturing setting. We emphasize that the expression of BirA* tagged lamins A and C were kept minimal (and almost undetectable) throughout our study as it is widely documented that overexpression of lamina proteins induce changes to the cellular state and, henceforth the cellular proteome, which will inevitably complicate downstream analysis of differential protein interactors. An extremely low expression of these proteins would allow the interrogation of differential interactions in an otherwise wildtype setting.

4.1 Materials and Methods

4.1.1 Materials

For Viral Production and Transduction

Expression plasmid for BioID fusion protein

Lentiviral packaging plasmids (Δ8.9 and vSVG)

HEK 293T viral packaging cells

100 mm tissue culture dishes Fugene6 (Promega)

DMEM (Gibco, Cat No. 11995-065)

Phosphate-buffered saline

Opti-MEM Growth Medium (see recipe) 10 mg/mL

Polybrene Beckman Ultracentrifuge SW28 rotor

Ultracentrifuge tubes

For SILAC labelling

SILAC growth media (see recipe)

SDS PAGE buffer (see recipe)

SDS-PAGE electrophoresis unit

For LARGE scale Cell Culture and Immunoprecipitation

SILAC growth medium (see recipe)

0.05% Trypsin, EDTA (Gibco)

Phosphate-buffered saline

Cell lysis (Hypotonic buffer)

BioID lysis buffer (see recipe)

Biotin

50 mM Tris·Cl, pH 7.4 Streptavidin agarose beads (Solulink, Cat No. N1000)

Control Agarose Resin (Peirce, Cat No. 26150)

Wash buffer 1 (see recipe)

Wash buffer 2 (see recipe)

Wash buffer 3 (see recipe)

1× elution buffer (see recipe)

DNase/RNase-free tubes, 15 ml and 50ml conical tubes.

Sonicator (Branson Sonifier-250 or equivalent)

Rotator Bradford Assay Protein

Quantification Kit and plate reader

SDS-PAGE electrophoresis unit

Mass Spectrometry grade trypsin

4.1.2 Generation of Low expressing BioID Cell Lines Lentivirus Production.

This section describes the production of lentiviruses in 100mm dish format. For different size format, scale the reagents used proportionally to the surface area of the dish.

Day 0:

1. HEK 293T cells were seeded at a density of 2 x 10^5 cells/cm^2.

2. The cells were then incubated at 37^o C, 5% CO_2 overnight.

Day1:

1. HEK293T cells were transfected with packaging plasmids and a lentiviral vector that harbors the BirA-protein fusion gene.

2. Transfection was performed following the manufacturer's protocol for Fugene-6.

3. 33.8ul of fugene was added to 900 uL of serum free DMEM and incubated at room temperature for 5 min.

4. During this incubation, the plasmid mixture consisting of 3.75ug of Δ8.9, 2.5ug of vSVG, and 5ug of BirA-protein plasmid was prepared.

5. The plasmid mixture from step 4. is then added to the fugene solution from step 3.

6. The mixture from step 5. is then incubated at room temperature for 15 min at room temperature before adding to the 293T cells.

7. The cells are then incubated at 37^{o} C, 5% CO_2.

8. 3 hours post transfection, sodium butyrate was added at a final concentration of 10mM.

• Sodium butyrate is a histone deacetylase (HDAC) inhibitor. The supplementation of sodium butyrate will hence enhance global transcription and hence boost the production of lentiviruses.

Day2:

9. The transfection media is removed and the cells were washed with warm PBS to remove any trace of sodium butyrate. 10. 4.5 ml of Opti-MEM was added to the cells. The cells were then incubated at 37^{o} C, 5% CO_2 for 12 hours.

11. The supernatant was collected into 15 ml conical tubes and stored at 4 ^{o}C.

12. Steps 2 and 3 were then repeated for a total of 3 collections per dish.

13. The viruses were pooled and centrifuged at 500 rcf for 5 min for 3 times. Each time, the viral supernatant is transferred to a new conical tube.

• This step removes any cellular debris contributed by the HEK 293T cells.

14. The lentiviruses were then concentrated using a Beckman Ultracentrifuge at 25000

rpm for 2 hours at 4 ^{o}C.

15. Viral pellets were then resuspended at 10ul opti-MEM per ml of un-concentrated virus and stored at 4 ^{o}C overnight.

Day3:

16. Infection of cells for BioID. A serial dilution of the virus was used in the infection. This is done for two purposes. Firstly, we emphasize the need to obtain a cell line with minimal BirA-protein expression. Additionally, when studying differential interactomes of disparate proteins, it is worthwhile obtaining cell lines that express almost the same amounts of the Bir-A protein across the experiment.

17. Infections are always performed In the presence of polybrene at a final concentration of 10ul/ml.

Day4:

18. Viral supernatant was removed and the cells were incubated with growth media at 37^{o} C, 5% CO_2.

Day5:

19. Selection drugs were added to ensure that the majority of the population expresses the BirA-protein.

Once the cell lines have been established, they have to be validated for correct

expression and localization. Additionally, the enzymatic activity of BirA should not be compromised when incorporated into the physiological environment relevant to the protein of interest. The latter can be validated several ways such as performing a small scale immunoprecipitation, post biotin incubation, followed by a western blot probing for biotinylated proteins via Horse Radish Peroxidase (HRP) conjugated Streptavidin. In the

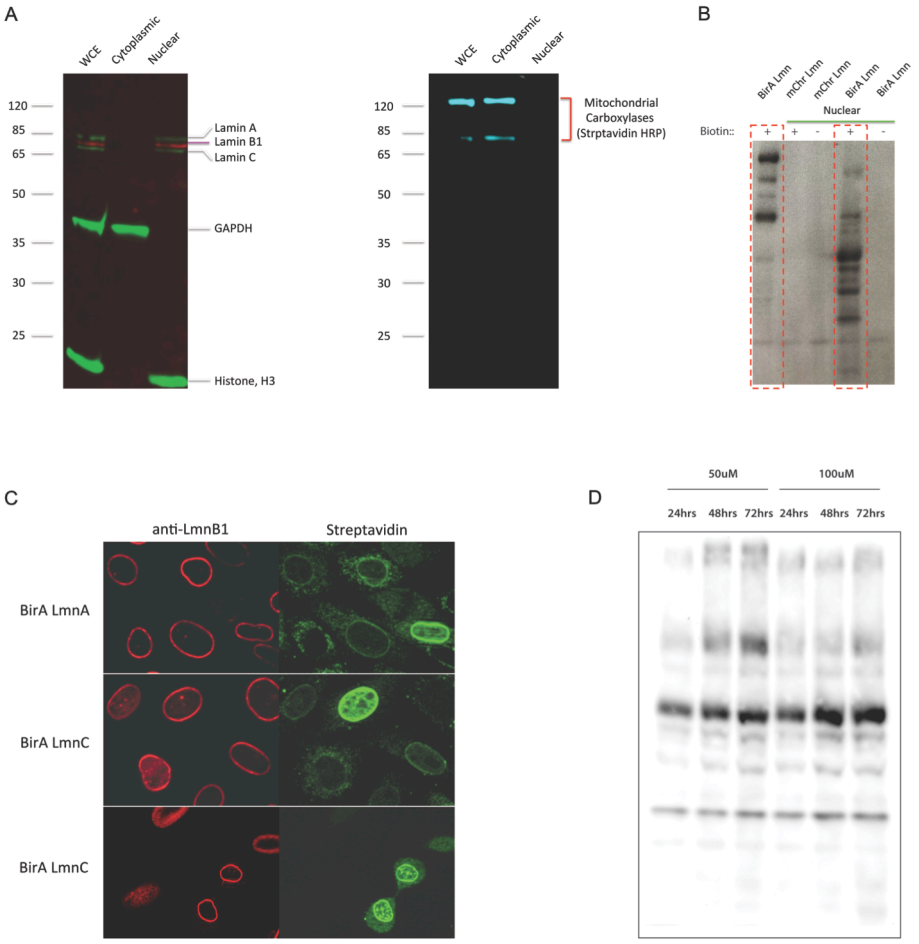


Figure 4.1. Optimization of the BioID protocol. (A) Left: western blots showing presence of nuclear proteins (lamins A, B1 and C and histone H3) but undetectable cytoplasmic traces, by GAPDH (green, 37 kDa) in the nuclear lysates. Right: Undetectable endogenously biotinylated carboxylases in nuclear lysates (Cyan; probed by Streptavidin-HRP). (B) Post streptavidin affinity purification western blot probed with streptavidin-HRP. Boxed in red are direct comparisons of the regular BioID protocol with affinity purification following isolation of nuclear lysates. (C) Representative images of BirA-LmnA or BirA-LmnC expressing cells incubated with biotin stained with antibodies against lamin B1 (red) and Cy2 conjugated streptavidin to detect biotinylated proteins (green). (D) Post streptavidin affinity purification western blot probed with streptavidin-HRP for testing various concentrations of biotin (50 uM and 100 uM) and incubation times (24, 48, 72 hours).

original BioID protocol, it was proposed that whole cell extracts are sufficient for visualization of biotinylated proteins by western blot. However, we note that this is challenging with low expression levels of BirA-protein, as is needed to effectively test lamin protein interactions without perturbing nuclear morphology and function (Figure 4.1). We therefore designed a protocol to specifically enrich for nuclear proteins and show that our simple nuclear prep prevents contamination of endogenously biotinylated mitochondrial proteins (Figure 4.1). As an alternative method to validate appropriate biotinylation of target proteins, cells can be fixed and subjected to an immunofluorescence staining protocol with fluorophore conjugated streptavidin. The fluorescence intensity for cells expressing BirA would be greater than control cells that have no BirA expression. In some cases, the fluorescence signal is seen to co-localize with cellular structures relevant to the protein of interest, validating the effectiveness of this technique. In our case, biotinylated proteins are seen to reside at the nuclear periphery as evident from a nuclear "rim" staining (Figure 4.1).

4.1.3 SILAC labeling of cell lines

In this protocol, we describe a two state SILAC experiment to identify differential interatomes of A type lamin variants, namely lamin A, lamin C. Cell lines expressing the different BirA tagged lamins are cultured in lysine and arginine depleted culture medium, supplemented with lysine (K) and arginine (R) of different isotopic masses. Depending on the experimental design, different combinations of isotopic labels can be chosen. In this chapter, the growth media for BirA-Lamin A and BirA-Lamin C cells were supplemented with K6R6 and K8R10 amino acids, respectively.

1. Cell lines were grown in small scale in their relevant growth media as described for at least 3 passages for maximal incorporation of the various isotopically labelled amino acids.

2. Whole cell extracts are generated for each line to check the percent incorporation of the isotopically labelled amino acids by mass spectrometry.

Once verified to have satisfactory incorporation, typically around 95%, the cultures are expanded for large scale harvest for immunoprecipitation. In this chapter, we describe a protocol for culturing the cells, in vivo biotin labelling and harvests in 150mm dishes. Importantly, the cells are always cultured in their respective SILAC media.

4.1.4 In vivo Labelling and Nuclear Enrichment

1. 72 hours prior to harvest, biotin is supplemented to the growth media at a final concentration of 50uM and incubated at 37 ^{o}C, 5% CO_2 for 24 hours.

2. The SILAC media is removed from the cells followed by washing of the cells with PBS. The cells are then trypsinised with 0.05% Trypsin, EDTA.

3. SILAC media is used to quench the trypsin. Cells were collected in 50 ml conical tubes and centrifuged at 700 rcf for 5 min.

4. The cell pellets were washed with ice cold PBS with protease inhibitors.

5. The cells were again centrifuged at 700 rcf for 5 min and resuspended with ice cold PBS with protease inhibitors. This step was repeated for a total of 3 PBS washes.

• These washes are important to remove biotin and trypsin prior to cell lysis.

6. The cells were again centrifuged at 700 rcf for 5 min.

7.The cell pellets were resuspended in ice cold cell lysis (hypotonic) buffer with protease inhibitors at approximately 50e6 cells/ml and incubated for at least 10 min on ice to obtain the cell nuclei.

8. Effectiveness of the lysis was verified by mixing a small aliquot of the lysis mix with trypan blue and visualizing under a light microscope. All cells should stain blue indicating complete cellular lysis.

9. The nuclei were centrifuged at 1500 rcf for 5 min and resuspended with ice cold PBS with protease inhibitors. This step was repeated for a total of 3 PBS washes.

10.The nuclei were again centrifuged at 1500 rcf for 5 min.

11. The nuclear pellets were lysed by adding BioID lysis buffer at 50e6 cells/ml.

• The lysis buffer in this step is devoid of DTT since DTT interferes with protein quantification which would be performed in the next section. We, however, add DTT back to the lysates post quantification.

12. Nuclear lysates can be frozen at -80 for storage or immediately processed for immunoprecipitation.

4.1.5 Processing and Quantification of Lysates

1. Triton-X was added to the lysates at a final concentration of 2%. Keep lysates on ice

during sonication.

• SDS precipitates at 4°C . A five-fold excess of Triton X-100 prevents SDS precipitation.

2. The lysates were sonicated using a Branson Sonifier 250 (or equivalent) at 30% duty cycle and an output level of 3 twice, followed by a constant pulse for 30s at an output level of 3. 2 min intervals were maintained between sonication cycles to prevent overheating.

• By this time, the lysates should not be viscous.

3. To provide for a more favorable immunoprecipitation condition, the lysates were mixed 1:1 with 50 mM Tris.CL pH 7.4.

4. To redissolve any precipitated proteins post addition of 50mM Tris.Cl, the lysates are sonicated twice at a constant pulse and an output level of 3 for 30s each.

5. The lysates are then centrifuged at 17,000g for 10 min and the supernatant was transferred to new tubes leaving behind a small insoluble pellet.

6. The lysates are then quantified by Bradford assay.

7. DTT was added to the lysates at a final concentration of 0.5 mM.

4.1.7 Immunoprecipitation

In this section, we describe the immunoprecipitation procedure for BioID. The protein input and amount of streptavidin beads have been optimized for our system. The

ratio of protein input to beads varies depending on the system and it is worthwhile setting up a pilot study for determining the amount of beads per mg of proteins into the immunoprecipitation. This is so that we can maximize the capture of biotinylated proteins while minimizing the amount of beads since streptavidin leeching or its proteolytic cleavage, during elution or on-bead trypsin digest respectively, would add to the interference to downstream mass spectrometry applications.

1. 5mg each of lysates collected from BirA-Lamin A and BirA-Lamin C expressing cells were pooled. The pooled lysates were precleared with agarose resin. 900uL of the resin removed of the storage buffer by centrifuging at 17,000g for 1 min. The resin was then washed with beads wash buffer and centrifuged at 17,000g for 1 min. The wash buffer was removed.

2. 1 ml of the combined lysates was used to resuspend the agarose resin pellet. This mix was then added to the entire 15mg mixed lysates and incubated at 4°C for 10min on a rocker.

3. During this incubation, the streptavidin beads are prepared. Similar to the preparation of the preclear agarose resin, 450uL of streptavidin conjugated beads were centrifuged at 5000g for 5 min and the storage buffer was removed and replaced with beads wash buffer. The beads remain in this buffer until the lysates are ready for streptavidin immunoprecipitation.

4. Centrifuge the lysates at 1500rcf for 5 min. Collect the supernatant in a new tube and discard the pelleted agarose resin.

• The lysates are now ready for immunoprecipitation.

5. Centrifuge the streptavidin beads that are still in wash buffer at 5000g for 5 min. Remove the wash buffer and resuspend the beads with 1 ml of the lysates.

6. The resuspended beads are then added to the entire 15 mg precleared nuclear SILAC labelled lysates for immunoprecipitation. This mixture was incubated on a rocker at 4°C overnight.

4.1.8 Post Immunoprecipitation Washes and Elution

All centrifugation and incubation steps in this section are performed at room temperature.

1. Centrifuge the lysates at 5000g for 5 min.

2. Remove the supernatant gently by pipetting. Try not to disturb beads pelleted at the bottom of the tube.

3. Gently tap the bottom of the tube to dislodge and partially resuspend the beads. Wash the beads with wash buffer 1 and then place the tube on a rotator for 8 min.

4. Remove the supernatant as described in steps 1 and 2 of this section.

5. Repeat step 3 once.

6. Remove the supernatant as described in steps 1 and 2 of this section.

7. Wash the beads with wash buffer 2 and then place the tube on a rotator for 8 min.

8. Again, remove the supernatant as described in steps 1 and 2 of this section.

9. Wash the beads with wash buffer 2 and then place the tube on a rotator for 8 min.

10. Remove the supernatant as described in steps 1 and 2 of this section.

11. Wash the beads with 50 mM Tris·Cl, pH 7.4 and then place the tube on a rotator for 8 min.

12. Repeat steps 10 and 11 for a total of two 50mM Tris.Cl washes.

13. Completely remove the supernatant as described in steps 1 and 2 of this section.

14. Add 150uL of elution buffer to resuspend the beads.

15. Heat the beads and elution buffer at 100^{o}C for 5 min.

16. Remove the beads and collect the eluate by centrifuging at 5000g for 5 min.

4.1.9 Post Immunoprecipitation Processing for Mass Spectrometry

1. The eluate was then run on an SDS PAGE gel. Multiple band excisions were performed on the entire lane that the lysates was run in.

2. In gel digestion was performed and digested peptides were retrieved.

3. Peptides were desalted by passing through a c18 column.

4.1.10 Recipe

Growth media:

500 ml DMEM (Gibco, Cat No. 11995-065) 50ml FBS 5 ml Penicillin/Streptomycin

5 ml L-glutamine

SILAC amino acid stocks

Arg0 or Arg6 or Arg10: approx. 84 mg/ml stock in PBS Lys0, Lys6 or Lys8: approx. 146 mg/ml stock in PBS

SILAC media:

Lysine and Arginine Free DMEM 50 ml FBS 5 ml Penicillin/Streptomycin Supplement with desired isotopically labelled lysine and arginine. (500ul each) **Cell Lysis buffer:**

5 mM PIPES pH 8.0 85 mM KCL 0.5% NP40 1× protease inhibitor (Protease Inhibitor Cocktail, Sigma Cat No. p8340-5ml; add just before use)

BioID Lysis buffer

50 mM Tris·Cl, pH 7.4

500 mM NaCl 0.2% SDS (w/v) Store up to 2 weeks at room temperature

1× protease inhibitor (Protease Inhibitor Cocktail, Sigma Cat No. p8340-5ml; add just before use)

Wash buffer 1:

2% SDS (w/v) Store up to 2 weeks at room temperature

Wash buffer 2:

0.1% (w/v) deoxycholic acid 1% (w/v) Triton X-100 1 mM EDTA 500 mM NaCl

50 mM HEPES, pH 7.5 Store up to 2 weeks at room temperature *Deoxycholic acid stock*

solution must be protected from light.

Wash buffer 3:

0.5% (w/v) deoxycholic acid 0.5% (w/v) NP-40 1 mM EDTA 250 mM LiCl

10 mM Tris·Cl, pH 7.4. Store up to 2 weeks at room temperature

Elution Buffer:

50 mM Tris·Cl, pH 6.8 12% sucrose 2% SDS 0.004% bromophenol blue 20 mM DTT (dithiothreitol)

4.2 Results and Discussion

4.2.1 Determination of differential interactors

We setup to determine differential protein interactomes between lamin A and lamin C by coupling BioID with SILAC-based mass spectrometry. Our scheme is highlighted in Figure 4.2. Briefly, we expressed BirA-LaminA at low levels in cells cultured in media containing 'medium' SILAC isotopes (Figure 4.2). In parallel, we expressed BirA-Lamin C at low levels in cells cultured in 'heavy' SILAC isotopes (Cambridge Isotope Laboratories). Nuclear lysates were prepared from each of these cell cultures (after an initial 2 week of culture followed by a label check to ensure incorporation of the heavy and medium isotopes and culture expansion in the respective media)., The lysates were quantified, equal amounts were mixed together and subjected to immunoprecipitation by streptavidin bound beads. The resulting proteins were eluted from the beads and run on an SDS-PAGE gel and 10 bands were isolated, extracted and

prepared for Mass Spectrometry analyses. From these experiments, we initially identified 201 differentially interacting proteins (Figure 4.2).

We highlight that, in our preliminary western bot data BirA* tagged lamins A or C exhibited equal labeling of lap2β. We therefore reasoned that, in our SILAC experiments, in which BirA* lamin C expressing cells were metabolically labeled with "heavy" amino acids and BirA* lamin A expressing cells were labeled with "medium" amino acids, it may be possible to use the lap2β ratio as an internal normalizer. We further show that the median ratio of heavy:light (BirA-Lamin C:BirA-LaminA) for our interactome study was 1.39 which was close to the interaction ratio observed for lap2β (1.43). Moreover, the expression of BirA* lamin C was about 1.4 times that of BirA* lamin A in the SILAC experiments (data not shown). All these data highly support the use of the lap2β ratio as an internal normalizer and a ratio of BirA-LaminC/BirA-LaminA of 1.4 as a 'zero' point

We therefore set a two fold threshold for proteins preferentially binding lamin C (>2.8 ratio) and for proteins preferentially binding lamin A (<0.7 ratio). Interestingly, proteins preferentially binding lamin C fell into two general classes: pre-mRNA processing and chromatin remodeling proteins suggesting that lamin C is important in both genome organization and the processing of mRNA transcripts.

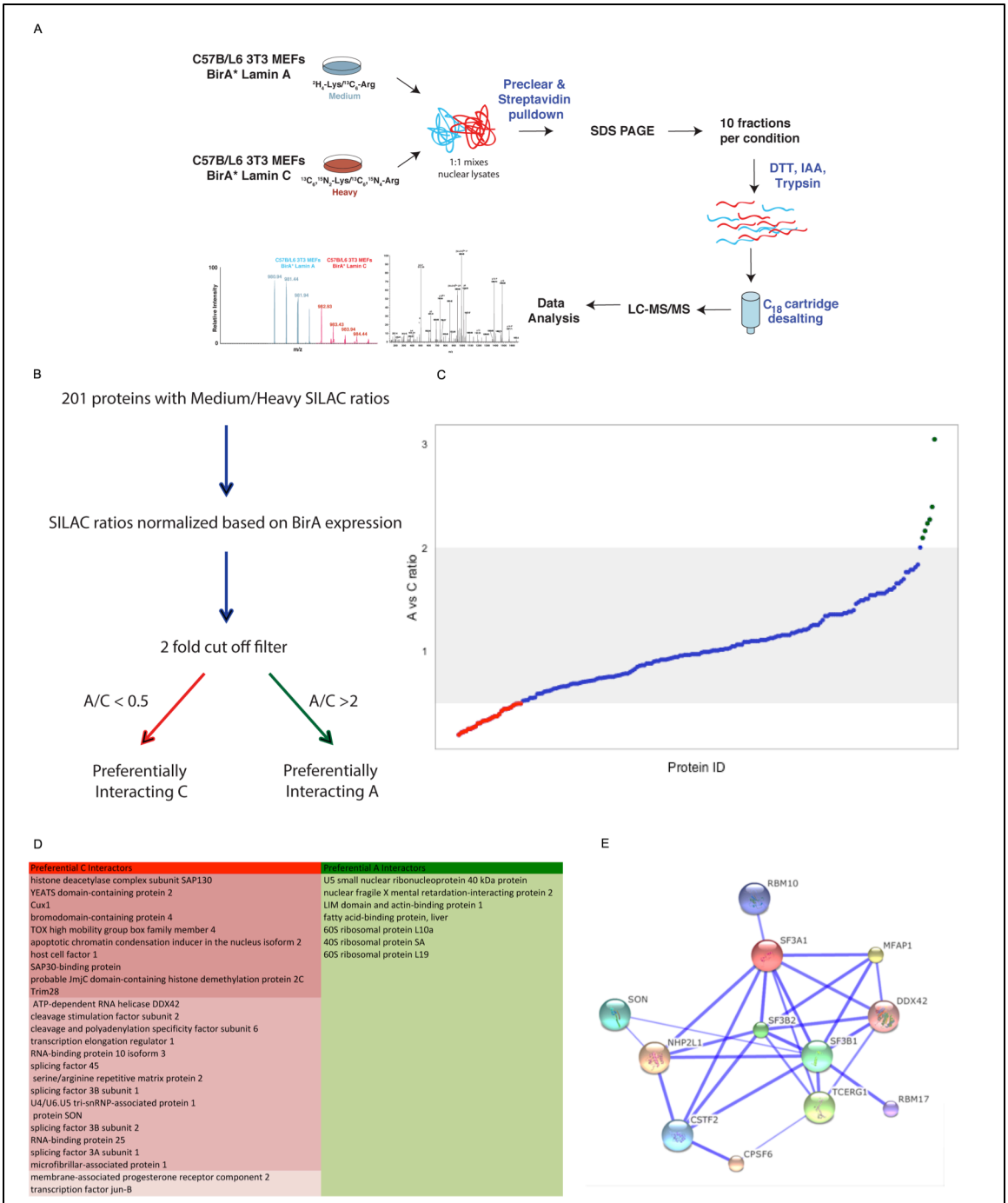


Figure 4.2. Experimental set up and analysis of BioID SILAC mass spectrometry experiment. (A) shows the general workflow for setting up the experiment. Cells expressing Bir*A lamin A or Bir*A lamin C were grown in medium and heavy media respectively and incubated with biotin X number of hours prior to harvest (number of hours determined by titration shown in Figure 4.1). Equal amounts of the harvested nuclear lysates are pooled, precleared with agarose beads and subjected to streptavidin conjugated agarose resin affinity purification. The eluate is then size fractionated by SDS PAGE. 10 fractions per lane were isolated and reduced, alkylated and digested to yield peptides which are then passed through c18 columns for desalting, prior to loading onto the mass spectrometry.

Figure 4.2. Experimental set up and analysis of BioID SILAC mass spectrometry experiment (continued) (B) Workflow for analyzing the resulting mass spectrometry data. SILAC ratios of Identified proteins were normalized based on the ratio of Bir*A lamin A to Bir*A lamin C expression. A cut off of SILAC ratio greater than 2 and less than 0.5 was set to determine differential protein interactors for lamin A and C, respectively. **(**C) shows the distribution of SILAC ratios obtained. Blue dots represent proteins that have SILAC scores within the 2 fold cut off. Green dots represent proteins with SILAC ratios of A:C greater than 2 and red dots represent proteins with SILAC ratios of A:C less than 0.5. (D) A list of differentially interacting proteins. Proteins interacting more preferentially with lamin C are highlighted in red and those interacting more preferentially with lamin A are highlighted in green. Different shades of red shows different functional classifications of lamin C interacting partners. (E) shows the string analysis for the class of pre-mRNA processing factors that preferentially interact with lamin C. The blue lines connecting each protein indicates the confidence of the interactions with thicker blue lines indicating a stronger confidence of interaction.

4.2.2 Lamin C specific Chromatin Remodelers

In the class of chromatin related factors, several of the proteins, TRIM28 which recruits the NuRD complex and SETDB1 and, SAP130 and SAP30 binding proteins that are subunits of the HDAC1 dependent Sin3a corepressor complex, are known to have transcriptionally repressive roles. Others, including Tox4, however, are associated with transcriptional activation and have been previously implicated in post mitotic chromatin structure control, among many diverse, yet unexplored functions.

While the contribution of these lamin C specific chromatin remodeling proteins will need to be elucidated in targeted studies, their preferential interaction with lamin C, but not lamin A, highlights the importance of lamin C in chromatin regulation.

We however hypothesize that, given the implications of some of these less well characterized chromatin remodeling proteins in mitosis, some of these lamin C specific chromatin remodelers may cooperate with lamin C in re-establishing genome organization post mitosis.

Protein Names	C/A Ratio	Normalized C/A Ratio
ATP-dependent RNA helicase DDX42	6.95	5
histone deacetylase complex subunit SAP130	6.3	4.532374101
YEATS domain-containing protein 2	6.03	4.338129496
Cux1	5.86	4.215827338
bromodomain-containing protein 4	5.37	3.863309353
TOX high mobility group box family member 4	5.32	3.827338129
cleavage stimulation factor subunit 2	5.1	3.669064748
apoptotic chromatin condensation inducer in the nucleus isoform 2	4.89	3.517985612
cleavage and polyadenylation specificity factor subunit 6	4.42	3.179856115
transcription elongation regulator 1	4.37	3.143884892
host cell factor 1	4.18	3.007194245
RNA-binding protein 10 isoform 3	4.13	2.971223022
membrane-associated progesterone receptor component 2	4.08	2.935251799
SAP30-binding protein	4.07	2.928057554
transcription factor jun-B	4.01	2.884892086
splicing factor 45	3.88	2.791366906
serine/arginine repetitive matrix protein 2	3.74	2.690647482
splicing factor 3B subunit 1	3.62	2.604316547
U4/U6.U5 tri-snRNP-associated protein 1	3.6	2.589928058
probable JmjC domain-containing histone demethylation protein 2C	3.42	2.460431655
protein SON	3.17	2.28057554
splicing factor 3B subunit 2	3.04	2.18705036
Trim28	3.03	2.179856115
RNA-binding protein 25	2.99	2.151079137
splicing factor 3A subunit 1	2.92	2.100719424
microfibrillar-associated protein 1	2.8	2.014388489
spliceosome-associated protein CWC15 homolog	2.73	1.964028777
caldesmon isoform X7	2.71	1.949640288
very low-density lipoprotein receptor isoform a precursor	2.71	1.949640288
zinc finger transcription factor Trps1	2.57	1.848920863
nuclear mitotic apparatus protein 1	2.55	1.834532374
tumor suppressor p53-binding protein 1	2.54	1.827338129
AT-rich interactive domain-containing protein 1B	2.44	1.755395683
actin, alpha cardiac muscle 1	2.44	1.755395683
splicing factor 1 isoform 1	2.42	1.741007194
heterogeneous nuclear ribonucleoprotein H2	2.28	1.64028777
Nup62#nuclear pore glycoprotein p62	2.27	1.633093525
Zbtb33, transcriptional regulator Kaiso	2.26	1.625899281
activity-dependent neuroprotector homeobox protein isoform 1	2.2	1.582733813
sentrin-specific protease 1	2.18	1.568345324
WW domain-containing oxidoreductase	2.17	1.561151079
SUN domain-containing protein 1	2.14	1.539568345
nuclear pore complex protein Nup98-Nup96	2.1	1.510791367
ruvB-like 1	2.09	1.503597122
14-3-3 protein sigma	2.06	1.482014388
U5 small nuclear ribonucleoprotein 200 kDa helicase	2.04	1.467625899
predicted gene 12657	2.03	1.460431655
tubulin alpha-1B chain	2.03	1.460431655
bcl-2-associated transcription factor 1 isoform 1	2.01	1.446043165
nucleoprotein TPR	1.97	1.417266187

Figure 4.3. Table of all proteins, with SILAC ratios, identified. Highlighted in yellow is the Lap2β internal normalizer.

emerin	1.95	1.402877698
translocon-associated protein subunit alpha	1.95	1.402877698
V-type proton ATPase catalytic subunit A	1.95	1.402877698
prelamin-A/C isoform A precursor	1.91	1.374100719
thyroid hormone receptor-associated protein 3	1.9	1.366906475
zinc finger CCCH domain-containing protein 14 isoform a	1.88	1.352517986
cleavage and polyadenylation specificity factor subunit 7 isoform 1	1.87	1.345323741
torsin-1A-interacting protein 1 isoform 1	1.86	1.338129496
targeting protein for Xklp2	1.83	1.316546763
tight junction protein ZO-1	1.81	1.302158273
filamin A-interacting protein 1-like isoform 1	1.81	1.302158273
nuclear pore complex protein Nup153	1.8	1.294964029
SUN domain-containing protein 2 isoform 1	1.79	1.287769784
neuroblast differentiation-associated protein AHNAK isoform X5	1.78	1.28057554
nuclear envelope pore membrane protein POM 121	1.78	1.28057554
SNW domain-containing protein 1	1.78	1.28057554
U2 small nuclear ribonucleoprotein A'	1.76	1.26618705
non-POU domain-containing octamer-binding protein	1.75	1.258992806
fibronectin isoform a precursor	1.74	1.251798561
latent-transforming growth factor beta-binding protein 1 isoform LTBP-1L precursor	1.72	1.237410072
DNA-directed RNA polymerase II subunit RPB1	1.71	1.230215827
nuclear pore complex protein Nup50	1.71	1.230215827
AT-rich interactive domain-containing protein 1A	1.68	1.208633094
YLP motif-containing protein 1	1.65	1.18705036
inner nuclear membrane protein Man1	1.62	1.165467626
ATP-dependent RNA helicase DDX3X	1.62	1.165467626
methylcrotonoyl-CoA carboxylase subunit alpha, mitochondrial	1.58	1.136690647
lamin-B receptor	1.57	1.129496403
serine/threonine-protein phosphatase PP1-alpha catalytic subunit	1.56	1.122302158
probable ATP-dependent RNA helicase DDX5	1.55	1.115107914
protein ELYS	1.55	1.115107914
heterogeneous nuclear ribonucleoprotein M isoform a	1.52	1.09352518
heterogeneous nuclear ribonucleoprotein A/B isoform 1	1.52	1.09352518
basement membrane-specific heparan sulfate proteoglycan core protein	1.52	1.09352518
procollagen-lysine,2-oxoglutarate 5-dioxygenase 3	1.49	1.071942446
SWI/SNF complex subunit SMARCC2	1.48	1.064748201
mitotic spindle assembly checkpoint protein MAD1	1.48	1.064748201
fibulin-2 isoform a precursor	1.46	1.050359712
pleiotropic regulator 1	1.46	1.050359712
propionyl-CoA carboxylase alpha chain, mitochondrial precursor	1.46	1.050359712
heterogeneous nuclear ribonucleoprotein K isoform 1	1.45	1.043165468
splicing factor 3B subunit 3	1.44	1.035971223
lamina-associated polypeptide 2 isoform beta	1.43	1.028776978
lamina-associated polypeptide 2 isoform alpha	1.42	1.021582734
alpha-actinin-1	1.42	1.021582734
heterogeneous nuclear ribonucleoprotein A1 isoform b	1.4	1.007194245
lamin-B2	1.4	1.007194245
lamin-B1	1.39	1
far upstream element-binding protein 2	1.39	1
annexin A2	1.39	1

Figure 4.3 (continued). Table of all proteins, with SILAC ratios, identified. Highlighted in yellow is the Lap2β internal normalizer.

heterogeneous nuclear ribonucleoprotein F	1.38	0.992805755
nucleoporin Nup43	1.38	0.992805755
histone H2A type 2-A	1.38	0.992805755
succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial precursor	1.38	0.992805755
poly(rC)-binding protein 2 isoform X2	1.36	0.978417266
heterogeneous nuclear ribonucleoprotein H	1.35	0.971223022
nestin	1.35	0.971223022
prohibitin	1.35	0.971223022
filamin-B isoform a	1.34	0.964028777
heterogeneous nuclear ribonucleoprotein D0	1.33	0.956834532
acetyl-CoA carboxylase 1 isoform X3	1.33	0.956834532
protein SCAF11	1.33	0.956834532
E3 SUMO-protein ligase RanBP2	1.32	0.949640288
poly(rC)-binding protein 1	1.32	0.949640288
ruvB-like 2	1.31	0.942446043
TAR DNA-binding protein 43 isoform 1	1.31	0.942446043
heterochromatin protein 1-binding protein 3 isoform X3	1.3	0.935251799
ATP synthase subunit beta, mitochondrial precursor	1.28	0.920863309
pre-mRNA-splicing factor RBM22	1.28	0.920863309
heterogeneous nuclear ribonucleoprotein U	1.26	0.90647482
phosphate carrier protein, mitochondrial	1.26	0.90647482
mitotic checkpoint protein BUB3	1.26	0.90647482
pre-mRNA-processing factor 40 homolog A	1.23	0.884892086
dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	1.23	0.884892086
nucleolar and coiled-body phosphoprotein 1 isoform C	1.22	0.877697842
actin-like protein 6A	1.22	0.877697842
drebrin	1.22	0.877697842
nucleophosmin isoform 1	1.22	0.877697842
protein GNAS isoform XLas	1.21	0.870503597
stress-70 protein, mitochondrial	1.21	0.870503597
histone H2B type 1-P isoform 1	1.21	0.870503597
eukaryotic translation initiation factor 2 subunit 1	1.2	0.863309353
filamin-A isoform X2	1.19	0.856115108
tropomyosin alpha-3 chain isoform Tpm3.1cy	1.19	0.856115108
pyruvate carboxylase, mitochondrial isoform 1	1.18	0.848920863
probable ATP-dependent RNA helicase DDX17 isoform 1	1.17	0.841726619
RNA-binding protein FUS	1.17	0.841726619
ADP/ATP translocase 1	1.16	0.834532374
low-density lipoprotein receptor isoform 1 precursor	1.16	0.834532374
heterogeneous nuclear ribonucleoprotein A3 isoform a	1.16	0.834532374
succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial precursor	1.16	0.834532374
histone H4	1.16	0.834532374
filamin-C	1.14	0.820143885
proliferating cell nuclear antigen	1.14	0.820143885
116 kDa U5 small nuclear ribonucleoprotein component isoform b	1.13	0.81294964
DAZ-associated protein 1	1.12	0.805755396
polypyrimidine tract-binding protein 3	1.12	0.805755396
actin, cytoplasmic 1	1.11	0.798561151
splicing factor, proline- and glutamine-rich	1.11	0.798561151
cytochrome b-c1 complex subunit 2, mitochondrial precursor	1.1	0.791366906
78 kDa glucose-regulated protein precursor	1.08	0.776978417

Figure 4.3 (continued). Table of all proteins, with SILAC ratios, identified. Highlighted in yellow is the Lap2β internal normalizer.

mitochondrial import receptor subunit TOM40 homolog	1.07	0.769784173
THO complex subunit 4	1.07	0.769784173
heterogeneous nuclear ribonucleoprotein L isoform X2	1.07	0.769784173
heterogeneous nuclear ribonucleoproteins A2/B1	1.05	0.755395683
calcium homeostasis endoplasmic reticulum protein	1.03	0.741007194
voltage-dependent anion-selective channel protein 1	1	0.71942446
chromosome alignment-maintaining phosphoprotein 1	1	0.71942446
ataxin-2-like protein isoform X4	0.99	0.712230216
cation-independent mannose-6-phosphate receptor precursor	0.99	0.712230216
vimentin	0.99	0.712230216
nucleolin	0.99	0.712230216
THO complex subunit 6 homolog	0.99	0.712230216
KH domain-containing, RNA-binding, signal transduction-associated protein 1	0.99	0.712230216
plectin isoform 12alpha	0.98	0.705035971
RNA-binding protein 14	0.97	0.697841727
RNA-binding protein EWS	0.97	0.697841727
nucleoporin Nup37	0.96	0.690647482
cytochrome b-c1 complex subunit 1, mitochondrial precursor	0.92	0.661870504
eukaryotic translation initiation factor 6	0.91	0.654676259
guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	0.9	0.647482014
prolow-density lipoprotein receptor-related protein 1 precursor	0.9	0.647482014
exosome complex component MTR3	0.89	0.64028777
voltage-dependent anion-selective channel protein 2	0.89	0.64028777
dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1 precursor	0.88	0.633093525
heat shock cognate 71 kDa protein	0.87	0.625899281
voltage-dependent anion-selective channel protein 3 isoform 1	0.87	0.625899281
pre-mRNA-processing factor 19 isoform 1	0.87	0.625899281
ATP synthase subunit alpha, mitochondrial precursor	0.86	0.618705036
elongation factor 1-alpha 1	0.86	0.618705036
cysteine and glycine-rich protein 2	0.86	0.618705036
heat shock protein HSP 90-alpha	0.86	0.618705036
cytoskeleton-associated protein 4	0.85	0.611510791
forkhead box protein C2	0.84	0.604316547
H/ACA ribonucleoprotein complex subunit 1	0.84	0.604316547
ankycorbin	0.83	0.597122302
ADP-ribosylation factor 2	0.81	0.582733813
transcription elongation factor SPT6	0.81	0.582733813
histone H1.3	0.8	0.575539568
histone H1.4	0.79	0.568345324
WD repeat-containing protein 43	0.76	0.54676259
eukaryotic initiation factor 4A-III	0.75	0.539568345
fragile X mental retardation syndrome-related protein 1 isoform 1	0.75	0.539568345
caprin-1 isoform a	0.74	0.532374101
alpha-actinin-4 isoform X4	0.73	0.525179856
U5 small nuclear ribonucleoprotein 40 kDa protein	0.67	0.482014388
nuclear fragile X mental retardation-interacting protein 2	0.64	0.460431655
LIM domain and actin-binding protein 1 isoform a	0.62	0.446043165
fatty acid-binding protein, liver	0.6	0.431654676
60S ribosomal protein L10a	0.59	0.424460432
40S ribosomal protein SA	0.56	0.402877698
60S ribosomal protein L19 isoform 1	0.44	0.316546763

Figure 4.3 (continued). Table of all proteins, with SILAC ratios, identified. Highlighted in yellow is the Lap2β internal normalizer.

Concluding Remarks

Recent evidence suggests that gene regulation can be influenced by nuclear architecture through large chromatin domain establishment as well as the local enrichment of regulatory and structural factors (Cremer et al., 2001, 2006; Elcock and Bridger, 2010; Fedorova and Zink, 2008; Ferrai et al., 2010; Misteli, 2005; Scaffidi and Misteli, 2006; Van Bortle and Corces, 2012). In particular, the nuclear periphery has been implicated in chromosome organization and gene regulation.

Molecular mapping of chromatin contacts at the nuclear periphery by DNA Adenine Methyltransferase (DamID, see chapter 2) has identified large chromatin domains termed Lmna associated domains (LADs: 0.1-10Mb) that display important cell type specific dynamics at the nuclear periphery. Specifically, genes within domains that gain association with the nuclear Lmna mostly become repressed and genes within domains that lost peripheral contact become activated or poised for expression, corroborating cytological studies correlating the proximity of specific gene loci relative to the nuclear Lmna to their transcriptional repression in several developmental systems (Kosak et al., 2002; Meister et al., 2010; Peric-Hupkes et al., 2010; Szczerbal et al., 2009; Williams et al., 2006; Yao et al., 2011). However, to date, very little data has been obtained on the role that lamins and INM proteins, the components of the nuclear periphery, play in LAD organization and function. We, therefore, sought to investigate the roles LmnA/C, one of the two main heterochromatin tethers at the nuclear periphery, play in LAD maintenance.

Interestingly, mutations in proteins of the nuclear periphery, in particular LMNA, result in developmental diseases that cover a spectrum of phenotypes ranging from tissue anomalies such as muscular dystrophies and lipodystrophies to premature aging (Bonne

et al., 1999; De Sandre-Giovannoli et al., 2002; Eriksson et al., 2003; Fatkin et al., 1999; Holt et al., 2001; Lin and Worman, 1993; Magracheva et al., 2009; Muchir et al., 2000; Novelli et al., 2002; Raffaele Di Barletta et al., 2000). These diseases have been collectively termed the Lmnopathies. At this point, it is unknown how these proteins, and in particular the lamins, regulate genome organization. Moreover, it is unknown if genome disorganization play a role in any of the over 500 diseases associated with mutations in proteins found in this region. Hitherto, most, if not all, studies on the roles LMN A/C play in normal developmental progression, in normal cell function, mostly investigate Lmn A and Lmn C collectively, with the underlying assumption that these two spliceoforms are functionally redundant (Al-Saaidi and Bross, 2015). However, our recent work investigating the mechanism of de novo LAD formation showed that cells depleted of both Lmn A and Lmn C, but not cells removed only of Lmn A, showed perturbed targeting to and maintenance of genic regions at the nuclear lamina (Harr et al., 2015).

This work, therefore, focuses on rigorously proving the functional non-redundancy of LmnA and LmnC in genome organization. Using a multi-prong approach, we sought to dissect the roles these highly identical spliceoforms may play in genome organization.

In this study, we showed that Lmn C, but not Lmn A, is required in maintaining LASes at the nuclear periphery. Additionally, we show that Lmn C is also important in regulating proper functional chromosome structure in interphase cells. These findings are consistent with the normal development of the Lmn C only mouse model that is devoid of Lmn A.

It is important to point out here that single cell techniques are important in detecting non-penetrant and asynchronized genome perturbations that cannot be picked up by the less sensitive population measures such as DamID. Nonetheless, population measures remain useful in characterizing the global architecture of specific cell types to gain a deeper understanding of architectural information that might be embedded within DNA sequences that may instruct genome organization. Moreover, by comparing global maps generated from disparate cell types, insights can be gained on the dynamics of genome reconfiguration during differentiation/development.

Because LmnC might be exerting its effect in interphase and/or through mitosis we further tested the relationship of LmnC to chromatin throughout the cell cycle, relative to LmnA and LmnB1. We find that LmnC lags behind both LmnA and B1 in its association with the periphery and remains in large puncta in the nucleoplasm into early G1. This non-peripheral localization of Lmn C correlated temporally with the reconfiguration of post mitotic genome to re-establish a functional interphase genome architecture (Chubb et al., 2002; Dileep et al., 2015; Dimitrova and Gilbert, 1999; Walter et al., 2003).

We next took a proteomics approach to identify differential interactors of LmnA and LmnC to gain a better understanding of how Lmn C might be informing genome organization. We show that Lmn C, but not Lmn A, preferentially binds chromatin remodeling proteins, highlighting its importance in regulating genome organization. Additionally, Lmn C also interacts with a subset of proteins, including Tox4, that have been documented to be involved in the governing of chromatin structure post mitotic exit

(Lee et al., 2010). All these data taken together, suggests that Lmn C, aided by Tox4 and other mitotic chromatin remodelers, might be involved in genome reconfiguration post mitosis.

In summary, we showed that LmnA and LmnC are non-functionally redundant. Specifically, we showed that LmnC, but not LmnA, is required for proper genome organization and this importance might be manifested, in part, through its involvement in post mitotic genome reorganization. We are currently in the pursuit of investigating whether the 6 amino acids unique to lamin C is necessary in its role in genome organization or if the lamin A specific tail is inhibitory in this process. One possibility that we can imaging is that the localization of lamin A might be regulated differently during mitosis so as to sequester lamin A away from chromatin. Additionally, we propose that variant specific functions exist and uncovering these specific roles Lmn A and Lmn C play in establishing proper cellular functions and their relative contributions to diseases will aid in the finer dissection of signaling/molecular pathways involved in the various laminopathies.

Appendices

AI: Functional Annotation of NETs Identified in liver, blood and muscle

Signaling	
Protein Name/Alternative Names	**Functions**
KIAA1161, NET37	IGF-II maturation [97]. Proposed lipid processing
LEMD2, NET25	Depletion causes extracellular signal-regulated kinase 1/2 hyperactivation which reduces myogenesis
ERLIN2, C8orf2, SPFH2, NET32	Proposed lipid microdomain organization
PPAPDC3, C9orf67, NET39	Antagonizes myogenesis via mTOR/IGF-II . Increases peripheral localization of chromosome 5, 11, 13. Decreased peripheral localization of chromosome 1
NCEH1, AADACL1, NET97	Ether lipid signaling pathway regulation in cancer
CTDNEP1, DULLARD, NET56	Removes phophoserine in Lipin 1 (lipid metabolism) . Antagonize BMP signaling required for neural induction
TMEM173, STING, NET23	Central signaling molecule essential to innate immune response upon infection by DNA pathogens
TMEM209, NET31	Nup205 stabilizing protein, increasing cMyc in nucleus
TMEM38A, TRIC-A, mNET1	Ca2+ channel
TMEM214	Microtubule association , Mediator of caspase-4 activation and apoptosis
KLHL31, mNET8	Transcriptional repressor of MAP/JNK pathway . Transcriptional repressor; overexpression inhibits SRE and TRE. Microtubule binding
CKAP4, CLIMP-63, ERGIC-63, P63, mNETY	Epithelial cell surface receptor for APF, Mediates the anchoring of the endoplasmic reticulum to microtubules.
SCARA5, NET33	Ferritin receptor. Downegulation associated with hepatocellular carcinoma; Inhibits FAK signaling pathway. Downregulated by Snail1 which leads to EMT-induced migration. Proposed innate immune activities.
APH1B, NET55	Aph-1b, a component of the γ-secretase enzyme complex that is involved in multiple (neuro)developmental signaling pathways.
RHBDD1, NET82	Serine protease important for cleaving BIK thereby modulating apoptosis mediated by BIK.
TMEM109,MG23, NET95	Purified MG23 behaved as a voltage-dependent, cation-conducting channel, permeable to both K(+) and Ca(2+). Protective role against UVC induced cell death.
LEMD3, MAN1	TGF/BMP modulation; sequestration of SMAD2/3.
EMD, LEMD5, STA	Senses force and activates downstream mechano-sensitive genes, influences centrosome positioning, responsible for scaffolding protein complexes at the periphery, may be involved in signal transduction/gene activation, binds and activates HDAC3, regulates signaling from the cell surface, may be involved in wnt signaling.
Genome Organization	
Protein Name/Alternative Names	**Functions**
PPAPDC3, C9orf67, NET39	Antagonizes myogenesis via mTOR/IGF-II [72]. Increases peripheral localization of chromosome 5, 11, 13. Decreased peripheral localization of chromosome 1.
MAGT1, OST3B, IAG2, NKP22	Chromatin condensation
STT3A, NET99	Increases peripheral localization of chromosome 5. Involved in ER glycosylation.
TAPBPL, NKP39	Increased peripheral localization of chromosome 5
TMEM201, NET5	Increase peripheral localization of chromosome 5. Cell cycle MT associating.
TMEM120A, NET29	Increased peripheral localization of chromosome 5, 13.
TM7SF2, NET47	Chromosome positioning (Chr 5, 11 to periphery, Chr 1,13 away from periphery). Regulates TNF-alpha, sterol reductase
LBR,DHCR14B, TDRD18	Cholesterol biosynthesis important to myeloid cell growth. Initiates and scaffolds heterochromatin. Roles in mitotic disassembly and re-assembly.
EMD, LEMD5, STA	Senses force and activates downstream mechano-sensitive genes, influences centrosome positioning, responsible for scaffolding protein complexes at the periphery, may be involved in signal transduction/gene activation, binds and activates HDAC3, regulates signaling from the cell surface, may be involved in wnt signaling.
TMPO, LEMD4, LAP2	Mediates transcriptional repression, involved in tethering chromatin to the nuclear periphery, involved in envelope breakdown and reassembly during mitosis
SUN1, UNC84A	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Important for nuclear rigidity and nuclear positioning. Binds telomeres in germ cells essential for homologous pairing during meiosis . Involved in DNA damage response.
SUN2, UNC84B	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks]. Important for nuclear rigidity and nuclear positioning. Binds telomeres in germ cells essential for homologous pairing during meiosis. Involved in DNA damage response.
SYNE1, C6orf98, Nesprin-1	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Involved in anchoring the nucleus beneath neuromuscular junction/synapse., Chromatin organization, NE architecture, vesicular transport, nuclear migration.
SYNE2, NUANCE, Nesprin-2	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Mechanotransduction, ciliogenesis, chromatin organization, NE architecture, nuclear migration.
Nuclear Structure	
Protein Name/Alternative Names	**Functions**
TMEM48, NDC1, NET3	NPC assembly, NE formation. Meiosis - synapse formation and resolution.
TMEM209, NET31	Nup205 stabilizing protein, increasing cMyc in nucleus
LBR, DHCR14B, TDRD18	Cholesterol biosynthesis important to myeloid cell growth. Initiates and scaffolds heterochromatin. Roles in mitotic disassembly and re-assembly
TMEM43, ARVD5, LUMA	Binds Emerin and influences its INM distribution. Also binds SUN2
EMD, LEMD5, STA	Senses force and activates downstream mechano-sensitive genes, influences centrosome positioning, responsible for scaffolding protein complexes at the periphery, may be involved in signal transduction/gene activation, binds and activates HDAC3, regulates signaling from the cell surface, may be involved in wnt signaling
TMPO, LEMD4, LAP2	Mediates transcriptional repression, involved in tethering chromatin to the nuclear periphery, involved in envelope breakdown and reassembly during mitosis
SUN1, UNC84A	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Important for nuclear rigidity and nuclear positioning. Binds telomeres in germ cells essential for homologous pairing during meiosis . Involved in DNA damage response.
SUN2, UNC84B	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Important for nuclear rigidity and nuclear positioning. Binds telomeres in germ cells essential for homologous pairing during meiosis . Involved in DNA damage response .
SYNE1, C6orf98, Nesprin-1	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Involved in anchoring the nucleus beneath neuromuscular junction/synapse., Chromatin organization, NE architecture, vesicular transport, nuclear migration.
SYNE2, NUANCE, Nesprin-2	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Mechanotransduction, ciliogenesis, chromatin organization, NE architecture, nuclear migration.
SYNE3, C14ORF49, Nesprin-3, Nesp3, NET53	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. NE-intermediate filament coupling.
TOR1AIP2, IFRG15, LULL1, NET9	Regulates the distribution and activity of TorA within the ER and NE lumen
POM121C	Required for reformation of nuclear envelop post mitosis in the presence of the Nup107-160 complex in vitro. Important for nucleating NPC assembly post mitosis (in some cases but not always) and during interphase).
NUP210	Important in early steps of NPC assembly and pore dilatation. This role has been challenged. Lumenal coiled coil region proposed to be important to fuse the ONM and INM at NPCs.

Figure AI.I. Table with full description of cellular functions of the NETs. The functions are classified into the categories of signaling, genome organization, nuclear structure, development and differentiation, lipid processing and others, with a handful of NETs having unknown functions.

Development/Differentiation	
Protein Name/Alternative Names	**Functions**
LEMD2, NET25	Depletion causes extracellular signal-regulated kinase 1/2 hyperactivation which reduces myogenesis .
PPAPDC3, C9orf67, NET39	Antagonizes myogenesis via mTOR/IGF-II. Increases peripheral localization of chromosome 5, 11, 13. Decreased peripheral localization of chromosome 1.
CTDNEP1, DULLARD, NET56	Removes phophoserine in Lipin 1 (lipid metabolism). Antagonize BMP signaling required for neural induction .
LEMD3, MAN1	TGF/BMP modulation; sequestration of SMAD2/3
TMEM14C, C6orf53, NET26	Required for Heme synthesis .
POPDC2, mNET3	Required for heart and skeletal muscle development
LBR,DHCR14B, TDRD18	Cholesterol biosynthesis important to myeloid cell growth. Initiates and scaffolds heterochromatin [59,62]. Roles in mitotic disassembly and re-assembly
APH1B, NET55	Aph-1b, a component of the γ-secretase enzyme complex that is involved in multiple (neuro)developmental signaling pathways.
EMD, LEMD5, STA	Senses force and activates downstream mechano-sensitive genes, influences centrosome positioning, responsible for scaffolding protein complexes at the periphery, may be involved in signal transduction/gene activation, binds and activates HDAC3, regulates signaling from the cell surface, may be involved in wnt signaling.
Lipid Processing	
Protein Name/Alternative Names	**Functions**
KIAA1161, NET37	IGF-II maturation. Proposed lipid processing
CTDNEP1, DULLARD, NET56	Removes phophoserine in Lipin 1 (lipid metabolism). Antagonize BMP signaling required for neural induction.
LPGAT1, FAM34A1, NET8	Lysophosphatidylglycerol acyltransferase activity
TM7SF2, NET47	Chromosome positioning (Chr 5, 11 to periphery, Chr 1,13 away from periphery). Regulates TNF-alpha, sterol reductase.
LBR,DHCR14B, TDRD18	Cholesterol biosynthesis important to myeloid cell growth. Initiates and scaffolds heterochromatin [59,62]. Roles in mitotic disassembly and re-assembly.
MBOAT5, LPCAT3, OACT5, mNET20	Lysophospholipid acyltransferase.
Other	
Protein Name/Alternative Names	**Functions**
STT3A, NET99	Increases peripheral localization of chromosome 5. Involved in ER glycosylation.
TMEM48, NDC1, NET3	NPC assembly, NE formation. Meiosis - synapse formation and resolution.
TMEM201, NET5	Increase peripheral localization of chromosome 5. Cell cycle MT associating
SCARA5, NET33	Ferritin receptor. Downegulation associated with hepatocellular carcinoma ; Inhibits FAK signaling pathway. Downregulated by Snail1 which leads to EMT-induced migration. Proposed innate immune activities.
TMEM214	Microtubule association, Mediator of caspase-4 activation and apoptosis
NRM, Nurim	Apoptosis Suppression.
KLHL31, mNET8	Transcriptional repressor of MAP/JNK pathway. Transcriptional repressor; overexpression inhibits SRE and TRE. Microtubule binding
CKAP4, CLIMP-63, ERGIC-63, P63, mNETY	Epithelial cell surface receptor for APF, Mediates the anchoring of the endoplasmic reticulum to microtubules
VMA21, LOC20354, mNET4	Essential assembly chaperone of the vacuolar ATPase (V-ATPase)
TMEM161A, mNET10	Functional role in protection against oxidative stress
WDR33, NET14	Component of the cleavage and polyadenylation specificity factor (CPSF) module of the pre-mRNA 3' end processing complex
TMEM109,MG23, NET95	Purified MG23 behaved as a voltage-dependent, cation-conducting channel, permeable to both K(+) and Ca(2+). Protective role against UVC induced cell death.
TMEM70, mNET11, bNET80	Involved in mitochondrial ATP synthase biosynthesis.
TMEM53, NET4	Stress dependent cell cycle withdrawal.
TOR1AIP1, LAP1	Interacts with Emerin, essential role for skeletal muscle maintenance.
APMAP, C20orf3, NKP83	Arylesterase activity
WFS1, mNET2	Accumulates at the base of mitotic spindles during mitosis.
EMD, LEMD5, STA	Senses force and activates downstream mechano-sensitive genes, influences centrosome positioning, responsible for scaffolding protein complexes at the periphery, may be involved in signal transduction/gene activation, binds and activates HDAC3, regulates signaling from the cell surface, may be involved in wnt signaling.
SYNE1, C6orf98, Nesprin-1	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Involved in anchoring the nucleus beneath neuromuscular junction/synapse., Chromatin organization, NE architecture, vesicular transport, nuclear migration.
SYNE2, NUANCE, Nesprin-2	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Mechanotransduction, ciliogenesis, chromatin organization, NE architecture, nuclear migration.
SUN1, UNC84A	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Important for nuclear rigidity and nuclear positioning. Binds telomeres in germ cells essential for homologous pairing during meiosis. Involved in DNA damage response.
SUN2, UNC84B	Part of the LINC complex that bridges the nucleoskeletal to cytoskeletal Networks. Important for nuclear rigidity and nuclear positioning. Binds telomeres in germ cells essential for homologous pairing during meiosis. Involved in DNA damage response.
Unknown	
Protein Name/Alternative Names	**Functions**
C17orf32, TMEM199, NKP40	Unknown
TMEM41A, NKP91	Unknown
SEC11C, NKP13	Unknown
C17orf62, NKP9	Unknown
TMEM126A, NKP90, NET90	Unknown
METTL7A, NKP2	Unknown
C14orf1, ERG28, NET51	Unknown
FAM105A, NET20	Unknown

Figure AI.I. Table with full description of cellular functions of the NETs (continued). The functions are classified into the categories of signaling, genome organization, nuclear structure, development and differentiation, lipid processing and others, with a handful of NETs having unknown functions.

AII: DamID Seq scripts

Fastq Quality Trimmer:

```
#!/usr/bin/perl
use strict;
use warnings;

#run script like this perl script.pl infile.fastq outfile.fastq
sub getScoreDict {
	my %scores = ();
	$scores{"!"} = 0;
	$scores{"\""} = 1;
	$scores{"#"} = 2;
	$scores{"\$"} = 3;
	$scores{"%"} = 4;
	$scores{"&"} = 5;
	$scores{"'"} = 6;
	$scores{"("} = 7;
	$scores{")"} = 8;
	$scores{"*"} = 9;
	$scores{"+"} = 10;
	$scores{","} = 11;
	$scores{"-"} = 12;
	$scores{"."} = 13;
	$scores{"/"} = 14;
	$scores{"0"} = 15;
	$scores{"1"} = 16;
	$scores{"2"} = 17;
	$scores{"3"} = 18;
	$scores{"4"} = 19;
	$scores{"5"} = 20;
	$scores{"6"} = 21;
	$scores{"7"} = 22;
	$scores{"8"} = 23;
	$scores{"9"} = 24;
	$scores{":"} = 25;
	$scores{";"} = 26;
	$scores{"<"} = 27;
	$scores{"="} = 28;
	$scores{">"} = 29;
	$scores{"?"} = 30;
	$scores{"@"} = 31;
	$scores{"A"} = 32;
	$scores{"B"} = 33;
	$scores{"C"} = 34;
```

```
$scores{"D"} = 35;
$scores{"E"} = 36;
$scores{"F"} = 37;
$scores{"G"} = 38;
$scores{"H"} = 39;
$scores{"I"} = 40;
$scores{"J"} = 41;
$scores{"K"} = 42;
$scores{"L"} = 43;
$scores{"M"} = 44;
$scores{"N"} = 45;
$scores{"O"} = 46;
$scores{"P"} = 47;
$scores{"Q"} = 48;
$scores{"R"} = 49;
$scores{"S"} = 50;
$scores{"T"} = 51;
$scores{"U"} = 52;
$scores{"V"} = 53;
$scores{"W"} = 54;
$scores{"X"} = 55;
$scores{"Y"} = 56;
$scores{"Z"} = 57;
$scores{"["} = 58;
$scores{"\\"} = 59;
$scores{"]"} = 60;
$scores{"^"} = 61;
$scores{"_"} = 62;
$scores{"`"} = 63;
$scores{"a"} = 64;
$scores{"b"} = 65;
$scores{"c"} = 66;
$scores{"d"} = 67;
$scores{"e"} = 68;
$scores{"f"} = 69;
$scores{"g"} = 70;
$scores{"h"} = 71;
$scores{"i"} = 72;
$scores{"j"} = 73;
$scores{"k"} = 74;
$scores{"l"} = 75;
$scores{"m"} = 76;
$scores{"n"} = 77;
$scores{"o"} = 78;
$scores{"p"} = 79;
$scores{"q"} = 80;
```

```
        $scores{"r"} = 81;
        $scores{"s"} = 82;
        $scores{"t"} = 83;
        $scores{"u"} = 84;
        $scores{"v"} = 85;
        $scores{"w"} = 86;
        $scores{"x"} = 87;
        $scores{"y"} = 88;
        $scores{"z"} = 89;
        $scores{"{"} = 90;
        $scores{"|"} = 91;
        $scores{"}"} = 92;
        $scores{"~"} = 93;

return \%scores;
}

sub qualTrim {
        #get reads reference
        my $ID = $_[0];
        my $read = $_[1];
        my $quals = $_[2];
        #get reference to the dictionary phred scores
        my %scoreList = %{$_[3]};

        my @read_seq_temp = split(",$read);
        my @quals_temp = split(", $quals);


        my $count = 0;
        my $mean_score = 0;
        my @flag;   #this stores how many times mean is <30

        #trim from 5' 1st
        while (defined($quals_temp[$count + 2]) && $mean_score < 30){
                $mean_score = ($scoreList{$quals_temp[$count]} +
                $scoreList{$quals_temp[$count+1]} +
                $scoreList{$quals_temp[$count+2]})/3;
                if ($mean_score < 30){
                        print "$mean_score\n";
                        push(@flag, $count);
                }
                print "$mean_score\n";
                $count++;
        }
        foreach my $k (@flag){
```

```
                my $first_read = shift(@read_seq_temp);
                my $first_qual = shift(@quals_temp);
        }

        #trim from 3' end
        $mean_score = 0;
        $count = scalar(@read_seq_temp)-1;
        @flag = ();  #empties it since used above
        while ($count >= 2 && $mean_score < 30){
                $mean_score = ($scoreList{$quals_temp[$count]} +
                $scoreList{$quals_temp[$count-1]} + $scoreList{$quals_temp[$count-
                2]})/3;
                if ($mean_score < 30){
                        push(@flag, $count);
                }
                $count--;
        }
        foreach my $k (@flag){
                pop(@read_seq_temp);
                pop(@quals_temp);
        }

        #need to get rid of seq with 2 nts left since the above loops don't look at
        remaining 2 elements
        if (scalar(@quals_temp)<3){
                return;
        }

        #generate trimmed reads and quals
        $quals = join('', @quals_temp);
        $read = join('', @read_seq_temp);

        print OUT "$ID\n$read\n+\n$quals\n";
}


my $start = time();
my $infile = $ARGV[0];
my $outfile = $ARGV[1];


open(IN, $infile) or die;
open (OUT, ">", $outfile) or die;

my %seq;
my $ID;
```

```
my $read;
my $quals;
my %score_;

#load quality scores
%score_ = %{getScoreDict()};
my $count = 0;
#load reads
while (my $i = <IN>){
	chomp $i;
	if ($count == 4){
		$count = 0;
	}
	if ($count == 0){
		$ID = $i;
	}
	elsif ($count == 1){
		$read = $i;
	}
	elsif ($count == 3){
		$quals = $i;
	   qualTrim($ID, $read, $quals, \%score_);
	}
	$count++;
}

#qualTrim(\%score_, \%seq);
my $end = time();

my $taken = $start - $end;
print "$taken\n";
```

Removing AdPCR

```
#!/usr/bin/perl
use strict;
use warnings;

#takes quality trimmed fastq files, remove contaminating sequences and regenerates fastq
file as input to bowtie #1
open (FH, "$ARGV[0]") or die;
open (OUT, ">$ARGV[1]") or die;

my $ID;
my $identifier;
my $line = 0;

while (my $i = <FH>){

        chomp ($i);

        if ($line == 4){
                $line = 0;

        }

        if($line == 0){
                $ID = $i;
                $line++;
                next;

        }

        if(($line ==  2) || ($line == 3)){
                $line++;
        next;
   }


        elsif($line == 1){

                $i =~
s/CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGA/\//g;
                $i =~
s/TCCTCGGCCGCGACCACGCTGCCCTATAGTGAGTCGTATTAG/\//g;
                $i =~ s/GATCCTCGGCCGCGACCGGTCGCGGCCGAGGATC/ /g;
```

```
		if ($i =~ /\//){

			$line++;
			next;

		}

		else {
			my @temp = split ('\s+',$i);
			my $count = 0;

			while ($temp[$count]){

				if (length($temp[$count])<25){

					$count++;
					next;

				}

				else {

					print OUT "$ID _ $count\n$temp[$count]\n+\n";
					print OUT "~"x length($temp[$count]);
					print OUT "\n";
					$count++;

				}
			}

		}
	}

	$line++;
}

close (OUT);
```

Trim 5' Ends

```
#!/usr/bin/perl
use strict;
use warnings;

#Takes unmapped reads from rm_bwt2_output [fastq] and remove 13 bases from 5' end
of reads

open (FH, "$ARGV[0]") or die;

open (OUT, ">$ARGV[1]") or die;

my $ID;
my $count = 0;

while (my $i = <FH>){
	chomp ($i);
	my $j = $i;

	if ($count == 4){
		$count = 0;
	}

	if ($count == 0){
		$ID = $i;
		$count++;
	}

	elsif ($count == 2 || $count == 3){
		$count++;
	}
	else {
		$i = substr($i, 13);
		if(length($i)>24){
			print OUT "$ID\n";
			print OUT "$i\n";
			print OUT "+\n";
			print OUT "~" x length($i);
			print OUT "\n";
		}
		$count++;
	}
}
```

Finding unmapped identifiers

```
use strict;
use warnings;

#takes the unmapped file from bowtie2 and looks for identifiers of unmapped reads.
#run like this: perl script_path bowtie2_unmapped_path output_identifiers_path

open (FH, "$ARGV[0]") or die;
open (OUT, ">$ARGV[1]") or die;

my $count = 0;


while (my $i = <FH>){
	chomp $i;
	if ($count ==4){
		$count = 0;
	}
	if ($count == 0){
		print OUT "$i\n";
		$count++;
	}
	else {
		$count++;
	}

}
```

Retrieving unmapped reads using the unmapped identifiers

```
#use strict;
#use warnings;

open (FH1, "$ARGV[0]") or die;    #stores IDs
open (FH2, "$ARGV[1]") or die;    #bowtie1 unmapped
open (OUT, ">$ARGV[2]") or die;

my %Wanted;

while (my $i = <FH1>){
        chomp $i;
        $Wanted{$i} = 1;
}

my $flag = 0;

while (my $j = <FH2>){
        chomp $j;


        if ($flag == 1){

                print OUT "$j\n";
                print OUT "+\n";
      print OUT "~" x length($j);
                print OUT "\n";
                $flag = 0;
        }


        if ($Wanted{$j}){

                print OUT "$j\n";
                $flag = 1;
        }
}
```

Trim 3' Ends

```
#!/usr/bin/perl
use strict;
use warnings;

#Takes unmapped reads from rm_bwt2_output [fastq] and remove 13 bases from 3' end
of reads
open (FH, "$ARGV[0]") or die;
open (OUT, ">$ARGV[1]") or die;

my $ID;
my $count = 0;

while (my $i = <FH>){
	chomp ($i);
	my $j = $i;

	if ($count == 4){
		$count = 0;
	}

	if ($count == 0){
		$ID = $i;
		$count++;
	}

	elsif ($count == 2 || $count == 3){

		$count++;
	}

	else {

		$i = substr($i, 0, -13);
		if(length($i)>24){
			print OUT "$ID\n";
			print OUT "$i\n";
			print OUT "+\n";
			print OUT "~" x length($i);
			print OUT "\n";
		}
		$count++;
	}

}
```

AIII: Visualization of genome wide LAD dynamics.

In chapter 3, we highlighted a striking temporal correlation of lamin C non-peripheral localization with post mitotic, large scale, re-establishment of genome architecture as previously documented. We hypothesize that lamin C may be informing this reconfiguration, in part, by informing the repositioning of LADs back to the nuclear periphery post mitosis (Figure AIII.1).

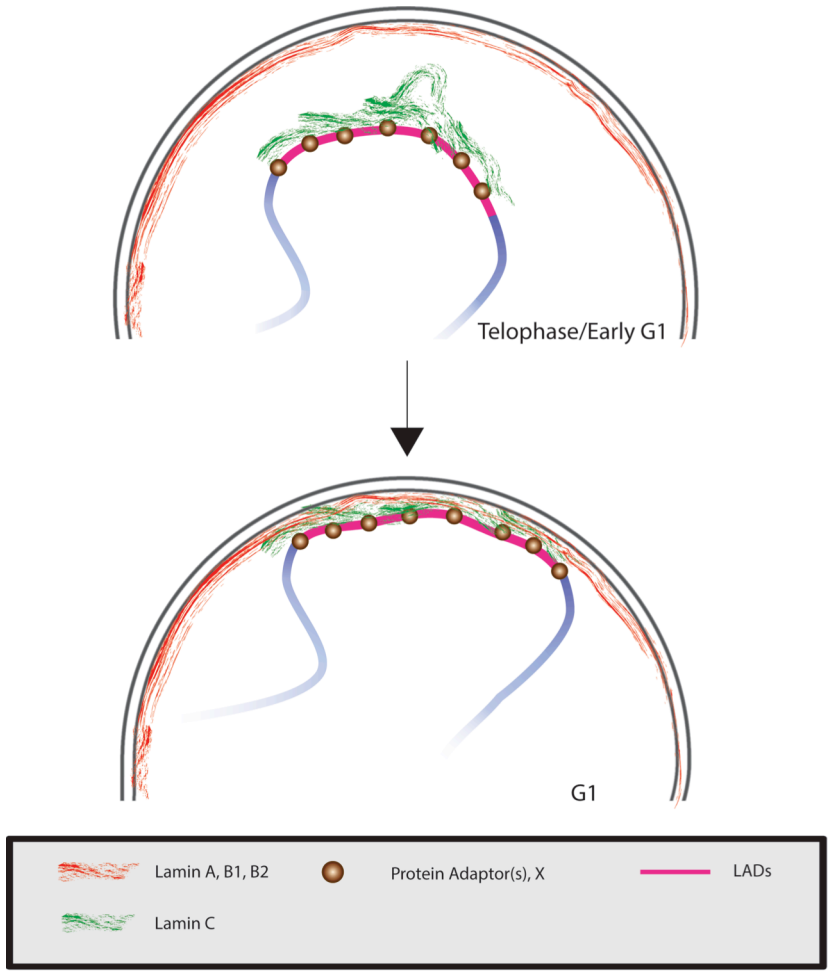


Figure AIII.1. Model of nucleoplasmic lamin C re-establishing LADs post mitosis. In telophase/early G1, nucleoplasmic lamin C binds to LADs (top) and "pulls" LADs towards the nuclear periphery where they remain in the next G1 phase (bottom).

To test this hypothesis, we re-engineered a recently published live cell imaging system for visualizing LADs. This system was originally comprised of 2 components, Dam-LmnB1 tagged at the N-terminus with a destabilizing domain (DD) and the catalytically inactive *Dpn*I enzyme tagged to GFP, named the ^{m6}A tracer (Kind et al., 2013). The ^{m6}A tracer retains the ability to bind Dam methylated DNA but has its endonuclease activity removed. The destabilization domain causes Dam-LmnB1 to be degraded by default. However, upon addition of the shield ligand that binds to the destabilization domain, Dam-LmnB1 becomes stably expressed and 'marks' DNA with ^{m6}A on the consensus sequence, GATCs, in close proximity. This then causes the GFP tagged ^{m6}A tracer to bind the Dam methylated DNA at the nuclear periphery, thereby allowing visualization of LADs in live cells.

To re-engineer this system, we synthesized the ^{m6}A tracer with overlapping oligonucleotides and tagged it downstream of the fluorescent protein, mCherry. Additionally, we highlight that the nuclear lamins become highly phosphorylated during mitosis and hence exhibit a diffused localization. To prevent random DNA methylation during mitosis, we tagged Dam-LmnB1 to the Cdt1 tag that has been used in the FUCCI system. The Cdt1 tag allows temporal control of Dam-LmnB1 expression through the cell cycle by allowing Dam-LmnB1 to be expressed highly in G1 phase and slowly becoming degraded in S phase so that the expression of Dam-LmnB1 in mitosis is minimal, if at all present (Sakaue-Sawano et al., 2008). This re-engineered system therefore allows the inducible 'marking' of interphase LADs that can then be followed through mitosis using the mCherry-tagged ^{m6}A tracer.

We validated the system by testing the ability of the, otherwise nucleoplasmic mCherry

^{m6}A tracer, to relocate to the nuclear periphery upon stabilizing DD-Dam-LmnB1 expression by adding the shield ligand (Figure AIII.2). Using this system, we performed pilot studies to trace the dynamics of LADs in interphase (in HEK 293, Figure AIII2 and C57B/L6 3T3 fibroblast cells, data not shown) and in mitosis (in HEK 293 cells) (Figure AIII2). While the interphase LADs are constantly juxtaposed at the nuclear periphery (and sometimes the nucleolar periphery), the LADs, together with the nuclear lamina, are highly dynamic and display 'synchronized' movements during interphase. During mitotic exit, the LADs can be seen to slowly reposition back into a 'rim' like structure indicating LAD re-establishment. However, the duration in which the process was captured was not long enough to capture the entire re-organization of the LADs post mitosis. This system will thus need to be validated further regarding the re-establishment of LADs post mitosis. In addition, we will use this system together with the expression of GFP tagged lamin C to observe the colocalization of lamin C with the LADs post mitosis, to gain insights to lamin C's potential involvement in post mitotic genome reconfiguration.

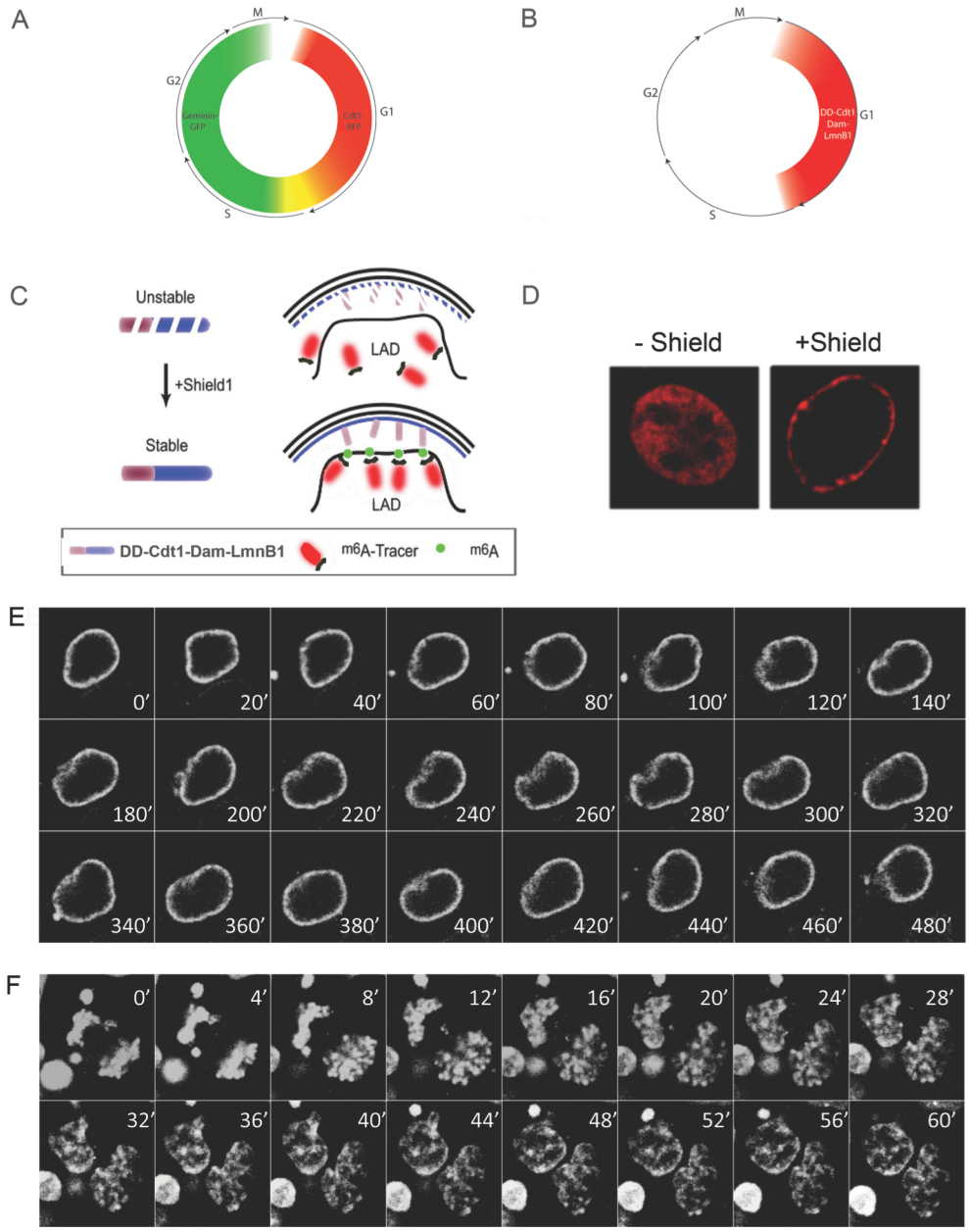


Figure AIII.2. Description of ^{m6}A system and validation. (A) Schematic of the FUCCI system. Cdt1-RFP is expressed in G1 phase and starts to get degraded in S phase when Geminin-GFP starts to get expressed (yellow). The expression of Geminin-GFP persists till mitosis before becoming degraded at telophase of mitosis. From telophase to the next early G1, both tagged proteins are not expressed (colorless). (B) Temporal expression pattern of Cdt1 tagged Dam-LmnB1. (C)) Schematic of the LAD-visualization system. The left shows the inducible proteotuner system - DD-Cdt1-Dam-LmnB1 is by default degraded unless bound to the shield stabilizing ligand, in which expression is restricted to G1 and early S. Right: mCherry ^{m6}A-tracer is nucleoplasmic in the absence of stable DD-Cdt1-Dam-LmnB1. The addition of shield stabilizes DD-Cdt1-Dam-LmnB1 expression and the presence of the ^{m6}A mark then sequesters the BirA mCherry ^{m6}A-tracer to LADs. (D) Validation of inducibility of the LAD-visualization system. Addition of shield causes sequestration of the ^{m6}A-tracer (red) to LADs. (E) Time lapse visualization of LADs in G1 at 20 min intervals. (F) Time lapse visualization of LADs through mitosis at 4 min intervals.

IV. BioLIPs - A novel technique for analyzing LADs interacting proteins

A growing body of evidence suggests that genome organization and sub-compartmentalization is important in the regulation of gene transcription programs. One obvious compartment, among others, that is important in providing a structural scaffold for the genome is the nuclear lamina and genomic regions that juxtapose at the lamina are known as LADs as previously mentioned. Having been widely documented to play important roles in development, attempts have been made to uncover characteristic features of LADs as well as how LADs become established and maintained. We have shown that the chromatin state (methylated histone H3 on K27 and K9) and cell-type specific factors play a role in establishing LADs but the connection between these DNA-bound interactors and the lamina remains elusive (Harr et al., 2015).

Histones are some of the most abundantly expressed proteins in the cell. Additionally, the enrichment of heterochromatin at the nuclear lamina suggests that a minimum appreciable abundance of histones would be detected in BioID experiments using the nuclear lamins as baits. However, our recent BioID mass spectrometry experiments, consistent with previous documentations, hardly detected any histone proteins (Roux et al., 2012, data not shown). These pieces of evidence led us to hypothesize that layers of information may lie between LADs/chromatin and the lamina (Figure AIV.1). While various flavors of genomic tools exist to unravel the different domains instructing genome architecture, hardly any proteomic means have been designed to study the proteomes of such domains. We introduce a novel system in this chapter for the analysis

of LAD proteomes, named BioID of LAD interacting proteins (BioLips). Although this chapter specifically describes the utility of this technique for studying LAD proteomes, it can be tweaked easily to understand the proteomic makeup of other chromatin domains.

BioLips is an adaptation of the ^{m6}A system as described in Chapter AIII, (shown in Figure AIV.1). Briefly, an interphase specific inducible expression of Dam-LmnB1 allows the marking of LADs in G1. The ^{m6}A tracer is then usurped in this system to target the biotin ligase mutant, used in BioID, to the LADs. This then allows for the inducible G1 specific tagging of LAD proteins with biotin for subsequent affinity purification and mass spectrometry analysis.

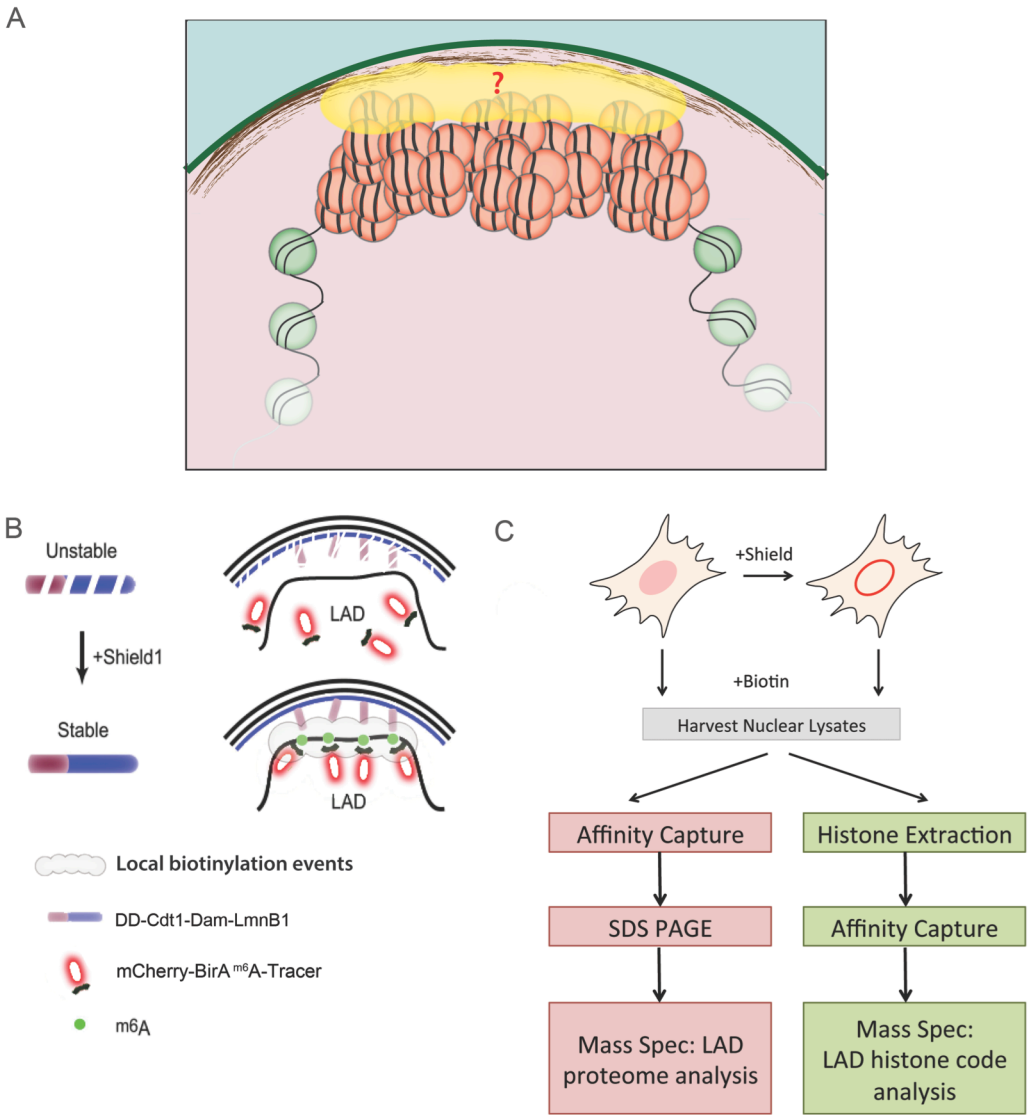


Figure AIV.1. The design of a novel system, BioLips, for interrogating LAD proteomes. (A) Cartoon of the LAD, lamina interface. LAD chromatin is shown in red signifying low or no transcription. Green non-LAD chromatin signifies detectable transcriptional output. A cloud between the LAD and the lamina indicates potential uncharacterized LAD proteins that links heterochromatin at the periphery to the lamina or INM proteins. (B) Schematic of the BioLips system. The left shows the inducible proteotuner system - DD-Dam-LmnB1 is by default degraded unless bound to the shield stabilizing ligand. Right: BirA mCherry ^{m6}A-tracer is nucleoplasmic in the absence of stable DD-Dam-LmnB1. The addition of shield stabilizes DD-Cdt1-Dam-LmnB1 expression and the presence of the ^{m6}A mark then sequesters the BirA mCherry ^{m6}A-tracer to LADs. This results in the local biotinlyation of LAD proteins. (C) Experimental workflow for the BioLips experiment. Cells pre- or post-treatment with shield are incubated with biotin followed by the harvest of nuclear lysates. The lysates are then either affinity purified for biotinylated proteins and mass fractionated by SDS PAGE or enriched for histones and affinity captured on streptavidin resin. The enriched material will then be analyzed on specific Mass Spectrometry setup for the analysis of LAD proteins and histone modifications, respectively.

References

Al-Saaidi, R., Bross, P., 2015. Do lamin A and lamin C have unique roles? Chromosoma 124, 1–12.

Bonne, G., Di Barletta, M.R., Varnous, S., Becane, H.M., Hammouda, E.H., Merlini, L., Muntoni, F., Greenberg, C.R., Gary, F., Urtizberea, J.A., Duboc, D., Fardeau, M., Toniolo, D., Schwartz, K., 1999. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. Nat Genet 21, 285–288.

Chubb, J.R., Boyle, S., Perry, P., Bickmore, W.A., 2002. Chromatin Motion Is Constrained by Association with Nuclear Compartments in Human Cells. Curr. Biol. 12, 439–445.

Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovei, I., Cremer, C., Cremer, T., 2001. Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. Chromosome Res. 9, 541–67.

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., Fakan, S., 2006. Chromosome territories--a functional nuclear landscape. Curr. Opin. Cell Biol. 18, 307–16.

De Sandre-Giovannoli, A., Chaouch, M., Kozlov, S., Vallat, J.-M., Tazir, M., Kassouri, N., Szepetowski, P., Hammadouche, T., Vandenberghe, A., Stewart, C.L., Grid, D., Lévy, N., 2002. Homozygous defects in LMNA, encoding lamin A/C nuclear-envelope proteins, cause autosomal recessive axonal neuropathy in human (Charcot-Marie-Tooth disorder type 2) and mouse. Am. J. Hum. Genet. 70, 726–36.

Dileep, V., Ay, F., Sima, J., Vera, D.L., Noble, W.S., Gilbert, D.M., 2015. Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication-timing program. Genome Res.

25, 1104–13.

Dimitrova, D.S., Gilbert, D.M., 1999. The spatial position and replication timing of chromosomal domains are both established in early G1 phase. Mol. Cell 4, 983–93.

Elcock, L.S., Bridger, J.M., 2010. Exploring the relationship between interphase gene positioning, transcriptional regulation and the nuclear matrix. Biochem. Soc. Trans. 38, 263–7.

Eriksson, M., Brown, W.T., Gordon, L.B., Glynn, M.W., Singer, J., Scott, L., Erdos, M.R., Robbins, C.M., Moses, T.Y., Berglund, P., Dutra, A., Pak, E., Durkin, S., Csoka, A.B., Boehnke, M., Glover, T.W., Collins, F.S., 2003. Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. Nature 423, 293–8.

Fatkin, D., MacRae, C., Sasaki, T., Wolff, M.R., Porcu, M., Frenneaux, M., Atherton, J., Vidaillet, H.J., Spudich, S., De Girolami, U., Seidman, J.G., Seidman, C., Muntoni, F., Müehle, G., Johnson, W., McDonough, B., 1999. Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction-system disease. N. Engl. J. Med. 341, 1715–24.

Fedorova, E., Zink, D., 2008. Nuclear architecture and gene regulation. Biochim. Biophys. Acta 1783, 2174–84.

Ferrai, C., de Castro, I.J., Lavitas, L., Chotalia, M., Pombo, A., 2010. Gene positioning. Cold Spring Harb. Perspect. Biol. 2, a000588.

Harr, J.C., Luperchio, T.R., Wong, X., Cohen, E., Wheelan, S.J., Reddy, K.L., 2015. Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins. J. Cell Biol. 208, 33–52.

Holt, I., Clements, L., Manilal, S., Brown, S.C., Morris, G.E., 2001. The R482Q lamin A/C mutation that causes lipodystrophy does not prevent nuclear targeting of lamin A in adipocytes or its interaction with emerin. Eur. J. Hum. Genet. 9, 204–8.

Kosak, S.T., Skok, J.A., Medina, K.L., Riblet, R., Le Beau, M.M., Fisher, A.G., Singh, H., 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. Science 296, 158–62. doi:10.1126/science.1068768

Lee, J.-H., You, J., Dobrota, E., Skalnik, D.G., 2010. Identification and characterization of a novel human PP1 phosphatase complex. J. Biol. Chem. 285, 24466–76.

Lin, F., Worman, H., 1993. Structural organization of the human gene encoding nuclear lamin A and nuclear lamin C. J. Biol. Chem. 268, 16321–16326.

Magracheva, E., Kozlov, S., Stewart, C.L., Wlodawer, A., Zdanov, A., 2009. Structure of the lamin A/C R482W mutant responsible for dominant familial partial lipodystrophy (FPLD). Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 65, 665–70.

Meister, P., Towbin, B.D., Pike, B.L., Ponti, A., Gasser, S.M., 2010. The spatial dynamics of tissue-specific promoters during C . elegans development. Genes Dev 24, 766–782. doi:10.1101/gad.559610.of

Misteli, T., 2005. Concepts in nuclear architecture. Bioessays 27, 477–87.

Muchir, A., Bonne, G., van der Kooi, A.J., van Meegen, M., Baas, F., Bolhuis, P.A., de Visser, M., Schwartz, K., 2000. Identification of mutations in the gene encoding lamins A/C in autosomal dominant limb girdle muscular dystrophy with atrioventricular conduction disturbances (LGMD1B). Hum. Mol. Genet. 9, 1453–9.

Novelli, G., Muchir, A., Sangiuolo, F., Helbling-Leclerc, A., D'Apice, M.R., Massart, C.,

Capon, F., Sbraccia, P., Federici, M., Lauro, R., Tudisco, C., Pallotta, R., Scarano, G., Dallapiccola, B., Merlini, L., Bonne, G., 2002. Mandibuloacral dysplasia is caused by a mutation in LMNA-encoding lamin A/C. Am. J. Hum. Genet. 71, 426–31.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S.W.M., Solovei, I., Brugman, W., Gräf, S., Flicek, P., Kerkhoven, R.M., van Lohuizen, M., Reinders, M., Wessels, L., van Steensel, B., 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol. Cell 38, 603–13. doi:10.1016/j.molcel.2010.03.016

Raffaele Di Barletta, M., Ricci, E., Galluzzi, G., Tonali, P., Mora, M., Morandi, L., Romorini, A., Voit, T., Orstavik, K.H., Merlini, L., Trevisan, C., Biancalana, V., Housmanowa-Petrusewicz, I., Bione, S., Ricotti, R., Schwartz, K., Bonne, G., Toniolo, D., 2000. Different mutations in the LMNA gene cause autosomal dominant and autosomal recessive Emery-Dreifuss muscular dystrophy. Am. J. Hum. Genet. 66, 1407–1412. doi:10.1086/302869

Scaffidi, P., Misteli, T., 2006. Lamin A-dependent nuclear defects in human aging. Science 312, 1059–63.

Szczerbal, I., Foster, H. a, Bridger, J.M., 2009. The spatial repositioning of adipogenesis genes is correlated with their expression status in a porcine mesenchymal stem cell adipogenesis model system. Chromosoma 118, 647–63. doi:10.1007/s00412-009-0225-5

Van Bortle, K., Corces, V.G., 2012. Nuclear organization and genome function. Annu. Rev. Cell Dev. Biol. 28, 163–87.

Venkatraman, E. S., & Olshen, A. B. (2007). A faster circular binary segmentation algorithm

for the analysis of array CGH data. Bioinformatics (Oxford, England), 23(6), 657–63.

Walter, J., Schermelleh, L., Cremer, M., Tashiro, S., Cremer, T., 2003. Chromosome order in HeLa cells changes during mitosis and early G1, but is stably maintained during subsequent interphase stages. J. Cell Biol. 160, 685–97.

Williams, R.R.E., Azuara, V., Perry, P., Sauer, S., Dvorkina, M., Jørgensen, H., Roix, J., McQueen, P., Misteli, T., Merkenschlager, M., Fisher, A.G., 2006. Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. J. Cell Sci. 119, 132–40.

Yao, J., Fetter, R.D., Hu, P., Betzig, E., Tjian, R., 2011. Subnuclear segregation of genes and core promoter factors in myogenesis. Genes Dev. 25, 569–580. doi:10.1101/gad.2021411.

Xianrong Wong

Johns Hopkins **xwong2@jhmi.edu**
School of Medicine (410) 982-9987
Rangos 580.17
855 North Wolfe Street
Baltimore, MD 21205

EDUCATION

Johns Hopkins School of Medicine, Baltimore, MD 2009-2015
Ph.D. Biochemistry, Cellular and Molecular Biology, Department of Biological Chemistry

National University of Singapore, Singapore 2004-2008
B. Eng. Summa Cum Laude

University of California, San Diego 2007
Student's Exchange Program, toward completion of B.Eng.

RESEARCH EXPERIENCE

Johns Hopkins School of Medicine, Baltimore, MD (2009-2015)
Center for Epigenetics and Department of Medicine
Ph.D. Mentor: Dr. Karen L. Reddy

- Made highly sought after tools to study the differential roles lamin A and C play in genome organization.
- Discovered that lamin C, and not lamin A, is unique in its role in genome organization and found that lamin C may be aiding (with other accessory proteins) genome re-configuration post mitosis, as one potential avenue for instructing genome organization.
- Developed a novel tool to interrogate the proteomes of lamina associated domains.

Institute of Medical Biology, Singapore (2008-2009)
Mentor: Colin L. Stewart

- Made induced pluripotent stem cells (iPSCs) from HGPS patient cells for the potential screening of drug targets.

National University of Singapore, Singapore (2007-2008)
Mentor: James Goh, Cho Hong and Siew Lok Toh

- Optimized a novel approach to tissue engineering bone grafts for spinal fusion surgery

Institute of High Performance Computing, Singapore (2006)
Mentor: Rosales C. Fernandez

- Investigated various dielectrophoretic single-cell trap designs using Boundary Element Method.

ACADEMIC EMPLOYMENT

Research Specialist **2008-2009**
Institute of Medical Biology, Singapore

Research was carried out under the direction of Colin Stewart.
- Made induced pluripotent stem cells (iPSCs) from HGPS patient cells for the potential screening of drug targets.

Research Intern **May/2006-Aug/2006**
Institute of High Performance Computing Research, Singapore

Research was carried out under the direction of Rosales C. Fernandez.
- Investigated various dielectrophoretic single-cell trap designs using Boundary Element Method.
- Submitted this work for publication.

PROFESSIONAL MEMBERSHIPS

American Society for Cell Biology

AWARDS, HONORS AND SCHOLORSHIPS

- National Science Scholarship (NSS) - Agency for Science and Technology, (2009-2014)
- Lee Kuan Yew Gold Medal: Best Graduating Student, 2008
- IES gold Medal: First in general proficiency throughout period of study, 2008

- Faculty of Engineering Innovative Researcher Award: High Achievement Award, 2008
- Provost's Honours for winter and spring quarters at University of California, San Diego. 2007
- Student's Exchange Program Award, 2007
- Dean's List for excellent academic achievement 2006, Semester 1
- Dean's List for excellent academic achievement 2006, Semester 2
- Dean's List for excellent academic achievement 2005, Semester 1
- NUS Donated Scholarship Award , 2005 Semerster 2
- Dean's List for excellent academic achievement , 2004, Semester 1

PUBLICATIONS

Invited / in preparation

Harr, J.C., **Wong X.**, Reddy K.L. (*invited, 2015*) **"**Targeting chromatin to the nuclear lamina." *Nucleus.*

Wong X*., Reddy K.L. (In preparation). An adaptation of DamID for Deep Sequencing.*

Wong X.*, Harr, J.C., Cutler, J., Pandey, A., Reddy, K.L. (In preparation) Differential Role of A type lamins in genome organization.*

Published

Wong, X., Reddy, K.L., 2015. Finding the Middlemen in Genome Organization. Dev Cell 35, 670–671.

Wong, X., Luperchio, T.R., Reddy, K.L. NET losses and gains: The role of the changing nuclear envelope proteome in genome regulation. Current Opinion Cell Biology 2014 May 30;28C:105-120

Luperchio, T.R, **Wong, X**., Reddy, K.L. Genome regulation at the peripheral zone: lamina associated domains in development and disease. Current Opinion Genes and Development Volume 25, April 2014, Pages 50–61

Harr, J.C., Luperchio, T.R., **Wong X.**, Cohen, E., Wheelan, S., Reddy, K. (2015) Directed targeting of chromatin to the nuclear lamina is mediated by chromatin state and A-type lamins". The Journal of Cell Biology 208(1) 33-52.
****This work is highlighted in a biobytes podcast, a Johns Hopkins SOM press release and in a Research Highlight (Nature Reviews Molecular Cell Biology 16, 68 (2015) doi:10.1038/nrm3948)***

Wong, X., Rosales, C., 2008. Robust dielectrophoretic single-cell trap design using BEM. Eng. Anal. Bound. Elem. 32, 388–394.

ABSTRACTS/PRESENTATIONS

Wong, X., Harr, J.C., Luperchio, T.R., Reddy, K.L. ABCs of genome organization: Differential roles for lamins A, B1 and C. Poster, Gordon Research Conference (Genome Architecture in Cell Fate & Disease), June 2015.

Wong, X., Harr, J.C., Luperchio, T.R., Reddy, K.L. The role of A type lamins in genome organization. Poster, ASCB, December 2014

Harr J.C.**, Wong X.,** Reddy, K. *Directed reorganization of chromatin to the nuclear lamina is mediated by chromatin state, YY1 and A-type lamins*. Talk, ASCB, December 2014

Harr, J.C**.**, Luperchio, T.R., **Wong X**., Cohen, E., Wheelan, S., Reddy, K. *Directed reorganization of chromatin to the nuclear lamina is mediated by YY1*. Abs, CSHL Nuclear Organization and Function, August 2014